

The association of active smoking with multiple cancers: national census-cancer registry cohorts with quantitative bias analysis

Tony Blakely · Jan J. Barendregt · Rachel H. Foster · Sarah Hill · June Atkinson · Diana Sarfati · Richard Edwards

Received: 24 October 2012 / Accepted: 27 March 2013 / Published online: 12 April 2013
© Springer Science+Business Media Dordrecht 2013

Abstract

Purposes (1) Determine the association of multiple cancers with smoking, focusing on cancers with an uncertain association; and (2) illustrate quantitative bias analysis as applied to registry data, to adjust for misclassification of smoking and residual confounding by alcohol and obesity.

Methods New Zealand 1981 and 1996 censuses, including smoking questions, were linked to cancer registry data giving 14.8 million person-years of follow-up. Rate ratios (RR) for current versus never smokers, adjusting for age, sex, ethnicity and socioeconomic factors were calculated and then subjected to quantitative bias analysis.

Results RR estimates for lung, larynx (including ear and nasosinus), and bladder cancers adjusted for measured confounders and exposure misclassification were 9.28 (95 % uncertainty interval 8.31–10.4), 6.14 (4.55–8.30), and 2.22 (1.94–2.55), respectively. Moderate associations were found for cervical (1.82; 1.51–2.20), kidney (1.29; 1.07–1.56), liver cancer (1.75; 1.37–2.24; European only), esophageal

(2.14; 1.73–2.65), oropharyngeal (2.30; 1.94–2.72), pancreatic (1.68; 1.44–1.96), and stomach cancers (1.42; 1.22–1.66). Protective associations were found for endometrial (0.67; 0.56–0.79) and melanoma (0.72; 0.65–0.81), and borderline association for thyroid (0.76; 0.58–1.00), colon (0.89; 0.81–0.98), and CML (0.66; 0.44–0.99). Remaining cancers had near null associations. Adjustment for residual confounding suggested little impact, except the RRs for endometrial, kidney, and esophageal cancers were slightly increased, and the oropharyngeal and liver (European/other) RRs were decreased.

Conclusions Our large study confirms the strong association of smoking with many cancers and strengthens the evidence for protective associations with thyroid cancer and melanoma. With large data sets, considering and adjusting for residual systematic error is as important as quantifying random error.

Keywords Smoking · Cancer incidence · Quantitative bias analysis · Epidemiological methods

Electronic supplementary material The online version of this article (doi:10.1007/s10552-013-0204-2) contains supplementary material, which is available to authorized users.

T. Blakely (✉) · R. H. Foster · J. Atkinson · D. Sarfati · R. Edwards
Department of Public Health, University of Otago,
PO Box 7343, Wellington, New Zealand
e-mail: tony.blakely@otago.ac.nz

J. J. Barendregt
Centre for Burden of Disease and Cost Effectiveness, School
of Population Health, University of Queensland, Brisbane,
Australia

S. Hill
Social Policy, School of Social and Political Science, University
of Edinburgh, Edinburgh, Scotland, UK

Introduction

The International Agency for Research on Cancer (IARC) and US Surgeon General (US-SG) now both recognize at least 12 cancers for which active smoking is a causal factor [1, 2]. The grouping of cancers by level of evidence and strength of association with smoking is shown in Table 1. More latterly, cancers with modest causal associations have been identified (e.g., nasal cavities and sinuses, squamous cell carcinoma of the cervix, renal cell carcinoma, and stomach cancer; Table 1). Only endometrial cancer in postmenopausal women clearly has a reduced rate among smokers. Consistently, null associations have been found for some common cancers (category 5 in Table 1). But

there remain a number of cancers for which there is uncertainty about any causal association with tobacco smoking (category 4 in Table 1). The substantive focus of this paper is on these latter ‘uncertain’ category 4 cancers.

Increasingly, epidemiological studies use large databases, often assembled through record linkage. Cancer registry studies—linked to mortality data—are often of this type. Record linkage beyond routine health data sets can enrich the data and extend possible analyses. In this paper, we use cancer registry data linked to national census data that happens to include smoking questions [3]. The periodic inclusion of a smoking question in New Zealand censuses (1976, 1981, 1996, 2006, and 2013) and subsequent linkage with health data allows rich epidemiological analyses (e.g., mortality and prevalence analyses we have previously published [4–8]). Linkage of 5 years of cancer registry data back to each of the 1981 and 1996 censuses creates some of the largest cohort studies of smoking and cancer incidence enabling a precise (i.e., narrow confidence interval) determination of relative risks of multiple cancers with smoking. But—in common with many registry and linkage studies—exposure assessment is not ideal (i.e., two simple questions on smoking status only) and data on some key potential confounders are missing (e.g., neither the NZ census nor cancer registry collects data on obesity or alcohol consumption). It is, however, possible to undertake quantitative bias analyses (QBA), in this case to adjust for possible misclassification bias of smoking and residual confounding by obesity and alcohol. (It is also possible that smoking contributes to lower BMI, and thus, adjusting for BMI is a form of overcontrol for confounding. However, for the purposes of this paper, we assume that the obesity is predominantly a confounder in that there are common causes of both smoking and increased BMI, or—expressed in directed acyclic graph terminology—BMI is on a backdoor path from smoking to cancer incidence.) Such QBA methods have existed in the epidemiological repertoire since 1959 [9], but are surprisingly infrequently used. Lash, Greenland, and colleagues have recently strongly argued for their wider application in epidemiology [10–12], noting that refusing to undertake QBA is tantamount to saying there is no remaining systematic error in one’s study. Thus, a second purpose of this paper is to demonstrate relatively simple QBA methods as applied to linked cancer registry data.

Brief overview of smoking–cancer associations

The association of breast cancer with smoking has been debated for sometime, with studies producing inconsistent findings [1, 2, 13–16]. For instance, while a major meta-analysis concluded there was little or no effect of smoking [13], two recent large cohort studies have found a modest

Table 1 Association of current smoking with cancers, and date of first pronouncement of any causal relationship by either the International Agency for Research on Cancer (IARC) or US Surgeon General (US-SG)

Category Cancer site	Strength of association ^a	Date of IARC or US-SG pronouncement, and comments ^b
<i>1. Causality established: moderate to strong association</i>		
Bladder	▲▲	IARC 1986 (bladder, ureter, and renal pelvis)
Larynx, nasal cavity, ear, and sinus		
Larynx	▲▲▲	US-SG 1964
Nasal cavity/sinus	▲▲	IARC 2004
Lip, mouth, and pharynx		
Oral cavity	▲▲	US-SG 1979
Pharynx	▲▲▲	IARC 1986
Lung, bronchus, and trachea	▲▲▲	US-SG 1964
Esophagus	▲▲	US-SG 1979 (extended to adenocarcinoma US-SG and IARC 2004)
<i>2. Causality established: weak to moderate association</i>		
Cervix	▲	US-SG 2004 and IARC 2004 (SCC only [1, 48])
Kidney (except renal pelvis)	▲	IARC 2004 and US-SG 2004
Pancreas	▲	IARC 1986
Stomach	▲	IARC 2004 and US-SG 2004
<i>3. Causality likely: protective association</i>		
Endometrium (postmenopausal)	▼	US-SG 2004
<i>4. Association and causality uncertain</i>		
Breast (female)	0 to ▲	IARC 2012: evidence of a small association. US-SG 2004: evidence suggestive of no causal association
Colorectum	0 to ▲	IARC 2012: evidence supports a causal association (although may be limited to certain subtypes). US-SG 2004: evidence suggestive, but not sufficient
Gallbladder and bile ducts	–	[As a rare cancer, lack of statistical power means any association is ‘unknown’ more than ‘likely null.’]
Leukemia	0	IARC 2004 (myeloid leukemia);
Myeloid leukemia	0 to ▲	US-SG 2004 (acute myeloid leukemia). [However, null in meta-analysis [47] of IARC data (1.09, 95 % CI 0.70–1.70), hence we include it in ‘causality uncertain.’]

Table 1 Continued

Category Cancer site	Strength of association ^a	Date of IARC or US-SG pronouncement, and comments ^b
Liver	0 to ▲	IARC 2004: evidence sufficient to infer a causal association. US-SG 2004: evidence suggestive, but not sufficient
Melanoma	0 to ▼	
Ovarian		IARC 2012: evidence sufficient to infer a causal association for mucinous ovarian cancer
Mucinous tumors	▲▲	
Other tumor types	0	
Thyroid	0 to ▼	IARC 2012: evidence suggestive of an inverse association
<i>5. Evidence suggestive of no association</i>		
Brain	0	
Hodgkin's	0	
Myeloma	0	
Non-Hodgkin's lymphoma (NHL)	0	[However, there has been a suggested association with follicular NHL.]
Prostate	0	
Testes	0	

Comments in square brackets are those of the authors of this paper, not IARC or US-SG

SCC squamous cell carcinoma

^a ▲▲▲ RR > 5; ▲▲ RR 2 to ≤5; ▲ RR 1 to ≤2; "0" is evidence suggestive of no causal association (i.e., RR = 1); "–" is evidence inconclusive or inadequate; ▼ significant protective relationship. Relative risk (or odds ratio) used in these groupings are as summarized by the IARC and US-SG reports [1, 2, 47, 49] and other meta-analyses [13, 18, 19, 21, 27, 30, 33, 43, 48, 50, 51]

^b US-SG reports relating to smoking and health (1964, 1979, 1982, 1990, 2004) are available at <http://www.surgeongeneral.gov/library/reports/index.html>, and IARC Monographs relating to smoking (volume 38: 1986; volume 83: 2004 and; volume 100E; 2012) are available at <http://monographs.iarc.fr/>

positive (increased risk) association of smoking with breast cancer [15, 16]. The latest IARC report (2012) identifies a consistent but weak positive association (RRs 1.1–1.3) [17].

The 2012 IARC report also states that current data support a causal relationship between smoking and colorectal cancer, although this may be limited to specific subtypes of colorectal cancer [17]. However, recent meta-analyses have reported weak positive associations with RRs of 1.07 (0.99–1.16) [18] and 1.17 (0.97–1.40) [19].

The association of smoking with leukemias and lymphomas is difficult to determine because they are heterogeneous groups of diseases, but there is some (inconsistent) evidence suggesting that the risk of follicular lymphoma may be increased [1, 20, 21]. IARC and the US-SG recognize a casual association of smoking with myeloid leukemias [1, 2, 22–24]. However, a meta-analysis of the

IARC data for myeloid leukemia did not find a statistically significant increase in risk [25].

Regarding ovarian cancer, three recent meta-analyses reported RRs of about two or more for mucinous ovarian tumors (10–20 % of all ovarian cancer cases) [26–28]. A causal association between smoking and mucinous, but not other types of, ovarian cancer has been identified by IARC [17].

A causal association of smoking with liver cancer is recognized by IARC [1], but not yet by the US-SG [2]. The strong correlation of smoking with other causes of liver cancer makes it difficult to confidently rule out residual confounding. Liver cancer is usually caused by chronic infection with hepatitis B or C viruses (HBV/HCV); HBV is particularly important in New Zealand among Māori and Pacific people [29]. Alcohol and other risk factors also make a contribution. A meta-analysis that carefully considered the confounding influence of alcohol, HBV, and HCV found RRs of 1.4–1.5 for liver cancer for current versus never smokers [30].

Melanoma is thought to have either no association or a possible negative association with smoking [1, 31, 32]. There is also some evidence of a negative association for thyroid cancer [33].

Methods

Data set

The 1981 and 1996 New Zealand censuses included questions on smoking. These two censuses and New Zealand Cancer Registry (NZCR) records for 1981–1986 and 1996–2001, respectively, were anonymously and probabilistically linked using geocodes, sex, date of birth, ethnicity, and country of birth (further details of methods are available elsewhere: www.uow.otago.ac.nz/cancertrends-info.html and [34]). The proportion of people with cancer who were successfully linked to a census record was 73 % for the 1981–1986 cohort and 80 % for the 1996–2001 cohort; accordingly, we calculated inverse probability of linkage weights for strata based on age, sex, ethnicity, residential mobility of the neighborhood, time since census, small area deprivation, cancer diagnosis, and region of the country. Assuming that there is negligible linkage variation by smoking status within strata of these covariates, these weights protect against linkage bias in final-weighted analyses.

Observations were included for all person-years of follow-up for current and never smokers aged 25 years and older. (Ex-smokers are excluded from the main analyses; however, results are available in Web Table 1 and Web Table 2).

Variables

The 1981 census included the following question “Tick the box which best describes your current cigarette smoking,” with three options: Never smoked cigarettes at all, or never smoked them regularly; do not smoke cigarettes now, but used to smoke them regularly (1 or more per day); and currently smoke cigarettes regularly (1 or more per day). The 1996 census used a two-step approach: (1) “Do you smoke cigarettes regularly (that is, one or more per day)?” (Count only tobacco cigarettes. Don’t count pipes, cigars, or cigarrillos); and for those answering “no,” (2) “Have you ever been a regular smoker of one or more cigarettes per day?”

Socio-demographic census data collected included age, sex, ethnicity, marital status, and multiple measures of socioeconomic position, including household income; highest educational qualification; household car access and housing tenure (both markers of asset wealth, and the latter also access to resources); and neighborhood deprivation (calculated for neighborhoods of about 100 people, using nine census variables [35]).

The outcome was a first cancer registered with the NZCR during the follow-up periods, excluding basal and squamous cell carcinomas of the skin and other in situ cancers. Cancers prior to 2000 were forward mapped from ICD-9 to ICD-10 codes. Site-specific analyses were conducted for adult cancer groupings (see Table 2 for groupings and ICD-10 codes). It was not possible to reliably identify the morphological types such as mucinous ovarian cancer.

Cohort analysis

Census respondents developing cancer were censored at the date of diagnosis. Unfortunately, due to privacy and logistical reasons, it was not possible to censor for death among people not developing a cancer or emigration out of New Zealand (estimated at less than 0.02 % per year; www.stats.govt.nz).

Age- and ethnicity-standardized (WHO world standard, further disaggregated by NZ ethnic distribution) incidence rates and rate ratios of cancer according to smoking status were calculated separately by sex and cohort. Three Poisson regression analyses were conducted for each cancer, for sexes and cohorts combined. First, a baseline regression model was run on the full data set, adjusting for sex, age, ethnicity, and cohort. Second, a model adjusting for the same variables was run on data restricted to those respondents with complete data on all socio-demographic factors. (About 26 % of respondents had missing data on one or more of the socioeconomic factors, usually household income which was classified as missing if any adult in the household had missing data due to either refusal or being away from their usual residence on census night.)

Comparing the second to baseline models allowed an assessment of selection bias, or at least that selection bias due to missing socioeconomic data. Third, a Poisson regression model adjusting additionally for socioeconomic factors was run on this restricted data set. Analyses by ethnic grouping were conducted for liver cancer only, due to the marked variation in the causal role of hepatitis B virus by ethnicity [29].

We selected ‘best’ rate ratio (RR) estimates from regression models 1 and 3 as follows. First, if there was negligible confounding by socioeconomic factors (which we conservatively defined as <3 % shift in the RR from model 2 to 3), then the best estimate was that from model 1. Second, if there was some confounding (i.e., ≥ 3 % shift in RR from model 2 to 3) and negligible selection bias (<3 % shift in RR from model 1 to 2), then the best estimate was that from model 3. Third, the remaining cancers had some evidence of *both* selection bias and confounding, and no direct ‘best’ estimate was available from our regression models. However, assuming that the observed amount of confounding moving from models 2 to 3 would have applied to the model 1 estimate if we actually had socioeconomic data on all census respondents, we were able to indirectly estimate what the ‘true’ RR estimate might have been. For example, if confounder adjustment resulted in a 15 % reduction in the RR from models 2 to 3, then our ‘derived’ estimate would be 85 % of the model 1 RR. (A worked example of these estimations is given in Box 1.)

Quantitative bias analysis: exposure misclassification and residual confounding

The census smoking variable is a crude measure, meaning our results were prone to exposure misclassification bias. We undertook quantitative bias analyses [10, 11] in purpose built Excel spread sheets using Monte Carlo simulation with Ersatz software (www.epigear.com). Details are provided in Web Appendices 1 and 2; briefly, a misclassification matrix of actual by self-reported smoking status was generated based on Heller et al. [36] and other studies and then modeled as a Dirichlet distribution applied to a subset of the crude data (45–64 and 65–84 year olds of European/other ethnicity from the 1996–2001 cohort, to avoid confounding by age or ethnicity) to generate corrected risk ratios. There was a near perfect linear association of the log of these adjusted RR estimates with the log of the crude RRs (Fig. 1), and we used this association to adjust all smoking–cancer RRs (Table 3).

The census does not include variables on potential confounders such as obesity and alcohol. We undertook bias analyses for cancers with known associations with either obesity or alcohol (detail in Web Appendix 1). Briefly, we used: (1) distributions of obesity and total

Table 2 Number of cancer events and age- and ethnicity-standardized incidence rate ratios (SRRs) for current versus never smokers aged 25+ years by sex and by cohort

Cancer group (ICD-10 code)	Males: 1981–1986		Males: 1996–2001		Females: 1981–1986		Females: 1996–2001	
	n (current/never)	SRR (95 % CI)	n (current/never)	SRR (95 % CI)	n (current/never)	SRR (95 % CI)	n (current/never)	SRR (95 % CI)
Bladder (C67)	432/231	1.89 (1.47–2.42)	393/525	1.86 (1.58–2.19)	120/204	2.07 (1.52–2.83)	138/390	2.25 (1.77–2.86)
Brain (C71)	117/165	0.73 (0.53–1.00)	99/264	0.71 (0.53–0.93)	72/159	1.05 (0.73–1.50)	63/234	0.97 (0.66–1.44)
Breast (C50)					1,317/3,507	0.91 (0.83–0.99)	1,689/5,871	0.92 (0.86–0.98)
Cervix (C53)					423/390	1.65 (1.34–2.02)	318/423	1.61 (1.33–1.95)
Colorectum (C18–20)	1,131/1,215	0.92 (0.83–1.03)	795/2,193	0.83 (0.75–0.92)	846/2,706	1.02 (0.92–1.13)	630/3,582	0.88 (0.80–0.97)
Colon (C18/C19)	687/771	0.90 (0.78–1.03)	489/1,458	0.78 (0.69–0.88)	594/1,935	1.02 (0.91–1.16)	453/2,700	0.84 (0.75–0.94)
Rectum (C20)	393/387	0.95 (0.78–1.15)	252/615	0.88 (0.74–1.06)	210/618	1.02 (0.83–1.27)	138/636	1.01 (0.80–1.27)
Endometrium (C54–5)					201/657	0.75 (0.59–0.94)	147/900	0.64 (0.50–0.82)
25–49 years					48/60	1.14 (0.70–1.84)	45/96	1.00 (0.62–1.61)
50 years+					153/600	0.68 (0.52–0.88)	105/804	0.58 (0.44–0.78)
Gallbladder/bile (C23–4)	48/36	1.74 (0.96–3.17)	30/54	1.00 (0.56–1.77)	36/138	0.97 (0.60–1.54)	21/123	0.75 (0.43–1.31)
Hodgkin's (C81)	63/51	1.28 (0.80–2.07)	36/63	1.11 (0.66–1.86)	27/69	1.00 (0.53–1.91)	21/66	0.89 (0.46–1.72)
Kidney (C64)	153/132	1.24 (0.84–1.83)	225/351	1.35 (1.10–1.65)	63/168	1.19 (0.80–1.76)	99/303	1.37 (1.04–1.80)
Larynx/ear/nasosinus (C30–2)	258/48	4.16 (2.21–7.85)	165/66	5.11 (3.55–7.35)	39/27	2.96 (1.16–7.55)	42/27	5.73 (3.19–10.28)
Leukemia (C91–5)	171/174	1.02 (0.75–1.38)	225/519	0.92 (0.76–1.11)	75/333	0.63 (0.43–0.92)	111/627	0.92 (0.72–1.18)
AML (C92.0)	51/48	1.53 (0.87–2.70)	69/117	1.09 (0.73–1.61)	21/117	0.47 (0.26–0.83)	45/183	0.99 (0.64–1.54)
CLL (C91.1)	54/60	0.79 (0.51–1.23)	102/252	0.94 (0.71–1.23)	21/108	0.80 (0.36–1.78)	36/276	0.83 (0.56–1.22)
CML (C92.1)	18/21	0.88 (0.39–1.98)	15/48	0.84 (0.45–1.55)	12/51	0.55 (0.22–1.38)	6/60	0.62 (0.26–1.51)
All non-AML	117/123	0.85 (0.60–1.21)	156/402	0.86 (0.69–1.08)	54/210	0.72 (0.45–1.16)	66/441	0.89 (0.66–1.19)
Liver (C22)	78/54	1.17 (0.74–1.87)	120/159	1.19 (0.89–1.60)	24/81	0.62 (0.30–1.26)	39/126	1.16 (0.75–1.82)
Lung, trachea, bronchi (C33–4)	2,577/432	5.13 (4.38–6.00)	1,923/588	7.31 (6.49–8.25)	813/366	6.36 (5.17–7.83)	1,272/684	7.65 (6.77–8.66)
Melanoma (C43)	321/528	0.62 (0.52–0.74)	405/1,740	0.54 (0.47–0.62)	414/1,179	0.81 (0.70–0.94)	465/2,313	0.73 (0.64–0.83)
Myeloma (C90)	105/114	1.03 (0.63–1.68)	84/246	0.72 (0.51–1.00)	54/195	0.82 (0.55–1.23)	51/303	0.93 (0.62–1.39)
NHL (C82–5)	186/219	0.84 (0.65–1.09)	228/561	0.89 (0.74–1.07)	108/324	0.99 (0.74–1.33)	147/777	0.88 (0.71–1.10)
Follicular lymphoma (C82)	33/36	0.84 (0.47–1.52)	42/102	0.85 (0.54–1.34)	30/57	1.57 (0.85–2.88)	30/144	0.81 (0.51–1.28)
Esophagus (C15)	159/96	1.60 (1.06–2.41)	159/159	2.48 (1.87–3.28)	57/147	1.98 (1.25–3.14)	60/198	1.89 (1.34–2.68)
Oropharynx (C00–14)	411/192	1.96 (1.55–2.48)	279/261	2.38 (1.92–2.96)	108/219	1.50 (1.07–2.09)	129/216	2.48 (1.90–3.25)
Ovary (C56)					180/534	0.80 (0.64–1.00)	234/768	1.05 (0.87–1.26)
Pancreas (C25)	231/144	1.34 (0.98–1.85)	174/258	1.31 (1.02–1.67)	108/297	1.23 (0.89–1.70)	138/456	1.49 (1.18–1.87)
Prostate (C61)	771/993	0.81 (0.71–0.93)	1,488/4,728	0.83 (0.77–0.89)				
Stomach (C16)	414/312	1.14 (0.89–1.46)	249/375	1.42 (1.16–1.73)	147/456	0.93 (0.70–1.23)	117/378	1.43 (1.07–1.91)

Table 2 continued

Cancer group (ICD-10 code)	Males: 1981–1986		Males: 1996–2001		Females: 1981–1986		Females: 1996–2001	
	n (current/never)	SRR (95 % CI)	n (current/never)	SRR (95 % CI)	n (current/never)	SRR (95 % CI)	n (current/never)	SRR (95 % CI)
Testicular (C62)	132/150	1.03 (0.75–1.42)	144/237	1.01 (0.76–1.33)				
25–44 years	111/135	0.96 (0.68–1.36)	126/198	0.95 (0.70–1.30)				
Thyroid (C73)	24/27	0.93 (0.34–2.56)	39/75	1.02 (0.59–1.75)	66/132	0.92 (0.58–1.47)	84/333	0.60 (0.42–0.84)

Person-years for the analysis of each cancer site was censored on the development of the first diagnosis of *that* cancer. Accordingly, person-years varied modestly across cancer sites, being the least for the most common cancers (e.g., lung, prostate, colorectal and breast) and the greatest for the rarer cancers (e.g., larynx and CML)—but varying by less than 0.5 %. Person-years ranged from: male current smokers—1,505,826–1,512,075 in 1981–1986 and 1,259,881–1,264,530 in 1996–2001; female current smokers—1,282,556–1,285,745 in 1981–1986 and 1,232,789–1,236,660 in 1996–2001; male never smokers—1,543,106–1,545,926 in 1981–1986 and 2,364,085–2,374,970 in 1996–2001; and female never smokers—2,488,021–2,496,549 in 1981–1986 and 3,121,801–3,135,797 in 1996–2001

AML acute myeloid leukemia, CML chronic myeloid leukemia, CLL chronic lymphocytic leukemia, NHL non-Hodgkin's lymphoma

alcohol consumption by smoking status among restricted age groups for European/other from the 2002/03 New Zealand Health Survey [37]; (2) the relative risk associations of obesity [38] and alcohol [39–42] with various cancers, and (3) uncertainty modeled with Dirichlet (prevalence data) and log normal distributions (relative risks).

Results

Smoking status was missing for 1.7 and 4.7 % of the 1981 and 1996 cohorts, respectively, giving 14.8 million person-years in total. Current smokers accounted for 36 % of person-years (41 % of the 1981–1986 cohort, and 31 % of the 1996–2001 cohort).

Table 2 shows the number of cancers for current and never smokers, and the age and ethnicity-standardized incidence rate ratios (SRRs). The SRRs were greater than 2.0, as expected, for larynx/ear/nasosinus and lung cancers.

Table 3 shows the RRs from Poisson regression models. Shifts in the RRs from model 1 to model 2 reflect selection bias arising from using the restricted data set of only respondents with complete data on all socio-demographic factors. Selection bias was notable for gallbladder and bile duct cancer (20 % increase in RR) and larynx/ear/nasosinus (16 % increase), and for liver (14 % increase; Māori/Pacific/Asian only) and lung/trachea/bronchi cancers (10 % increase).

Shifts in RRs from model 2 to model 3 (additionally adjusting for socioeconomic factors) reflect confounding. Strong confounding was apparent for melanoma (a 12 % increase in RR from 0.65 to 0.73), consistent with melanoma being more common among higher socioeconomic groups who are less likely to smoke. Rate ratios also increased modestly after adjustment for breast cancer, follicular lymphoma and prostate cancer. Modest 5–9 % reductions in the rate ratios after adjustment were evident for cervical, larynx/ear/nasosinus and liver cancers.

The third to last column of Table 3 presents 'best' empirical RR estimates from either model 1 or model 3 and 'derived' estimates for seven cancer groupings that had evidence of both confounding and selection biases (see "Methods" for details).

We also further adjusted for misclassification bias of the smoking exposure (details in Web Appendix 1). Briefly, we found that this adjustment had a negligible impact for our best estimate relative risks up to about 1.25, but for higher RRs, our best estimates increasingly underestimated the exposure misclassification-adjusted estimates. Notably, the association of the crude and exposure misclassification-adjusted risk ratios for 36 sex by age by cancer strata from the 1996–2001 cohort was near perfectly linear on the

Table 3 Poisson regression incidence rate ratios (RR) for current versus never smokers aged 25+ years, and ‘best’ RR allowing or adjusting for systematic error, sexes, and cohorts combined (numbers of cases and person-years for current and never smokers included in regressions are shown in Web Table 4)

Site	Model 1: RR (95 % CI)	Model 2: RR (95 % CI)	Model 3: RR (95 % CI)	Absolute (%) difference model 1–2 (selection bias)	Absolute (%) difference model 2–3 (confounding)	Best estimates (95 % uncertainty interval) after sequential adjustment for		Possible additional change correcting for alcohol and obesity confounding
						Selection bias and measured confounders ^a	Exposure misclassification ^b	
Bladder	2.01 (1.83–2.22)	2.07 (1.85–2.30)	2.11 (1.89–2.36)	0.06 (3 %)	0.04 (2 %)	2.01 (1.83–2.22)	2.22 (1.94–2.55)	NA
Brain	0.84 (0.72–0.97)	0.85 (0.72–1.01)	0.88 (0.74–1.05)	0.01 (1 %)	0.03 (4 %)	0.88 (0.74–1.05)	0.89 (0.72–1.10)	NA
Breast	0.94 (0.89–0.98)	0.95 (0.90–1.00)	1.00 (0.94–1.05)	0.01 (1 %)	0.05 (5 %)	1.00 (0.94–1.05)	1.03 (0.93–1.13)	+0.02 to 0.03
Cervix	1.82 (1.61–2.06)	1.89 (1.64–2.18)	1.74 (1.50–2.02)	0.07 (4 %)	–0.15 (–8 %)	1.68 (1.44–1.94)	1.82 (1.51–2.20)	NA
Colorectal (excl anal)	0.91 (0.87–0.96)	0.91 (0.87–0.96)	0.94 (0.89–0.99)	0 (0 %)	0.03 (3 %)	0.94 (0.89–0.99)	0.96 (0.87–1.05)	+0.01 to 0.02
Colon (C18/19)	0.88 (0.83–0.93)	0.88 (0.83–0.94)	0.90 (0.85–0.97)	0 (0 %)	0.02 (2 %)	0.88 (0.83–0.93)	0.89 (0.81–0.98)	
Rectum (C20)	0.98 (0.90–1.08)	0.99 (0.89–1.09)	1.00 (0.90–1.11)	0.01 (1 %)	0.01 (1 %)	0.98 (0.90–1.08)	1.00 (0.88–1.14)	
Endometrium 25 years+	0.68 (0.60–0.78)	0.66 (0.56–0.77)	0.66 (0.57–0.78)	–0.02 (–3 %)	0 (0 %)	0.68 (0.60–0.78)	0.67 (0.56–0.79)	+0.08
Endometrium 50 years+	0.62 (0.53–0.72)	0.61 (0.51–0.73)	0.62 (0.52–0.74)	–0.01 (–2 %)	0.01 (2 %)	0.62 (0.53–0.72)	0.60 (0.50–0.73)	
Gallbladder and bile ducts ^c	1.06 (0.84–1.35)	1.27 (0.97–1.67)	1.27 (0.96–1.67)	0.21 (20 %)	0 (0 %)	1.06 (0.84–1.35)	1.10 (0.83–1.45)	NA
Hodgkin’s	1.09 (0.85–1.41)	1.14 (0.85–1.53)	1.15 (0.85–1.55)	0.05 (5 %)	0.01 (1 %)	1.09 (0.85–1.41)	1.13 (0.84–1.51)	NA
Kidney	1.27 (1.12–1.44)	1.27 (1.10–1.46)	1.23 (1.06–1.42)	0 (0 %)	–0.04 (–3 %)	1.23 (1.06–1.42)	1.29 (1.07–1.56)	+0.07 to 0.10
Larynx, ear, nasosinus ^c	5.38 (4.35–6.66)	6.23 (4.83–8.04)	5.83 (4.50–7.56)	0.85 (16 %)	–0.40 (–6 %)	5.03 (3.88–6.53)	6.14 (4.55–8.30)	NA
Leukemia (all)	0.91 (0.82–1.02)	0.91 (0.80–1.04)	0.91 (0.80–1.04)	0 (0 %)	0 (0 %)	0.91 (0.82–1.02)	0.92 (0.80–1.07)	+0.02
AML	0.96 (0.78–1.18)	0.87 (0.69–1.10)	0.88 (0.69–1.11)	–0.09 (–9 %)	0.01 (1 %)	0.96 (0.78–1.18)	0.98 (0.77–1.26)	
CLL ^d	0.86 (0.71–1.03)	0.89 (0.72–1.09)	0.88 (0.71–1.10)	0.03 (3 %)	0.01 (1 %)	0.86 (0.71–1.03)	0.87 (0.69–1.09)	
CML	0.67 (0.47–0.97)	0.69 (0.45–1.05)	0.70 (0.46–1.08)	0.02 (3 %)	0.01 (1 %)	0.67 (0.47–0.97)	0.66 (0.44–0.99)	
All non-AML	0.89 (0.78–1.02)	0.93 (0.79–1.08)	0.93 (0.79–1.08)	0.04 (4 %)	0 (0 %)	0.89 (0.78–1.02)	0.90 (0.76–1.07)	
Liver ^e	1.21 (1.01–1.46)	1.32 (1.06–1.64)	1.23 (0.98–1.54)	0.11 (9 %)	–0.09 (–7 %)	1.13 (0.90–1.41)	1.18 (0.90–1.53)	
European/other 45 years+	1.65 (1.30–2.11)	1.66 (1.25–2.19)	1.62 (1.22–2.16)	0.01 (1 %)	–0.04 (–2 %)	1.65 (1.30–2.11)	1.75 (1.37–2.24)	–0.02 to –0.10
Māori/Pacific/ Asian	0.92 (0.69–1.22)	1.05 (0.73–1.50)	0.93 (0.64–1.34)	0.13 (14 %)	–0.12 (–11 %)	0.82 (0.57–1.18)	0.80 (0.53–1.20)	Unsure ^f
Lung, trachea, and bronchi	7.57 (7.14–8.03)	8.36 (7.80–8.96)	8.06 (7.51–8.65)	0.79 (10 %)	–0.30 (–4 %)	7.30 (6.80–7.83)	9.28 (8.31–10.4)	NA
Melanoma	0.66 (0.62–0.70)	0.65 (0.61–0.70)	0.73 (0.67–0.78)	–0.01 (–2 %)	0.08 (12 %)	0.73 (0.67–0.78)	0.72 (0.65–0.81)	NA
Myeloma	0.88 (0.75–1.03)	0.83 (0.69–1.01)	0.84 (0.70–1.02)	–0.05 (–6 %)	0.01 (1 %)	0.88 (0.75–1.03)	0.89 (0.73–1.09)	NA

Table 3 continued

Site	Model 1: RR (95 % CI)	Model 2: RR (95 % CI)	Model 3: RR (95 % CI)	Absolute (%) difference model 1–2 (selection bias)	Absolute (%) difference model 2–3 (confounding)	Best estimates (95 % uncertainty interval) after sequential adjustment for	Possible additional change correcting for alcohol and obesity confounding	
							Selection bias and measured confounders ^a	Exposure misclassification ^b
NHL	0.89 (0.80–0.99)	0.85 (0.75–0.95)	0.88 (0.78–1.00)	-0.04 (-4 %)	0.03 (4 %)	0.92 (0.81–1.04)	0.94 (0.79–1.10)	+0.02
Follicular lymphoma	0.92 (0.72–1.17)	0.88 (0.67–1.15)	0.95 (0.72–1.24)	-0.04 (-4 %)	0.07 (8 %)	0.99 (0.76–1.30)	1.02 (0.75–1.38)	NA
Esophagus	1.89 (1.62–2.19)	1.98 (1.67–2.35)	2.03 (1.71–2.42)	0.09 (5 %)	0.05 (3 %)	1.94 (1.63–2.31)	2.14 (1.73–2.65)	+0.10 to 0.20
Oropharynx	2.16 (1.93–2.42)	2.26 (1.99–2.57)	2.17 (1.91–2.48)	0.10 (5 %)	-0.09 (-4 %)	2.07 (1.82–2.36)	2.30 (1.94–2.72)	-0.40 to -0.10
Ovary	0.98 (0.85–1.11)	0.96 (0.83–1.11)	0.98 (0.84–1.14)	-0.02 (-2 %)	0.02 (2 %)	0.98 (0.85–1.11)	1.00 (0.84–1.19)	NA
Pancreas	1.56 (1.39–1.75)	1.62 (1.42–1.85)	1.64 (1.43–1.88)	0.06 (4 %)	0.02 (1 %)	1.56 (1.39–1.75)	1.68 (1.44–1.96)	NA
Prostate	0.81 (0.76–0.86)	0.82 (0.77–0.88)	0.88 (0.82–0.94)	0.01 (1 %)	0.06 (7 %)	0.88 (0.82–0.94)	0.89 (0.80–0.99)	NA
Stomach	1.37 (1.24–1.51)	1.39 (1.24–1.56)	1.34 (1.19–1.50)	0.02 (1 %)	-0.05 (-4 %)	1.34 (1.19–1.50)	1.42 (1.22–1.66)	NA
Testicular (25–44 years)	0.96 (0.77–1.19)	0.97 (0.76–1.23)	0.94 (0.73–1.21)	0.01 (1 %)	-0.03 (-3 %)	0.94 (0.73–1.21)	0.96 (0.72–1.28)	NA
Thyroid	0.74 (0.61–0.90)	0.73 (0.59–0.92)	0.76 (0.60–0.96)	-0.01 (-1 %)	0.03 (4 %)	0.76 (0.60–0.96)	0.76 (0.58–1.00)	NA

AML acute myeloid leukemia, CML chronic myeloid leukemia, CLL chronic lymphocytic leukemia, NHL non-Hodgkin's lymphoma, SES socioeconomic status

^a 'Best' estimates are from model 1 when there was no substantive confounding and from model 3 when there was confounding but no substantive selection bias. Estimates shown in italics are derived when there were both confounding and selection bias; see "Methods" for details of derivation

^b Exposure misclassification bias-adjusted 'best' estimates are those using the regression equation for the association of misclassification bias-adjusted estimates with crude estimates obtained from 36 quantitative bias adjustments on 45–64- and 65–84-year-old European/other strata (Fig. 1) and the average increase in total uncertainty compared to random error-only confidence interval. See text and Web Appendix 1 for details of method

Model 1 = adjusted for sex, age, ethnicity, and cohort for all data with these variables complete and smoking status known

Model 2 = adjusted for sex, age, ethnicity, and cohort for only data for which all of these variables plus all SES variables used in model 3 were complete and smoking status known

Model 3 = adjusted as for model 2 plus SES variables

Unless otherwise specified, age was specified in 5-year age groups and ethnicity was prioritized in the following order: Māori; Pacific; European/other

SES variables were as follows: car access (nil, 1, 2+ cars); education (no qualifications, school qualifications, post-school qualifications); tenure (not owned or free, owned); marital status (never, separated/divorced/widowed, currently married); household equivalized income (log); and neighborhood deprivation (quintile codes as continuous variable)

^c Age specified in 10-year age groups (5-year age group specification failed to see model converge)

^d Age specified in 10-year age groups, and ethnicity limited to European/other (to insure model convergence). Note that numbers of cancers and person-years are therefore just European/other
^e Due to the different causal factors for liver cancer by ethnicity (hepatitis B is responsible for 80 %+ of Māori and Pacific liver cancer, compared to 6 % or less for European/other [29]), smoking RRs are presented separately by ethnic groupings. Among European/other, numbers of cases aged 25–44 were too few for regression modeling, so a lower age limit of 45 was used—but 25 years retained for Māori/Pacific/Asian

^f Due to the heterogeneity by ethnicity in the RR association of smoking with liver cancer, and that our confounding quantitative bias analyses were conducted on European/other only, it is not possible to estimate the magnitude of confounding for all ethnic groups combined (even though European/other are by far the majority, Māori/Pacific/Asian rates of liver cancer are much higher) or Māori/Pacific/Asian specifically

Box 1 Demonstration of how the ‘best’ RR estimate was calculated for cancer of the larynx, ear, and nasosinus (as in Table 3)

Model 1 gives the RR for all current smokers with adjustment for sex, age, ethnicity, and cohort (i.e., 5.38)

In Model 2, we restrict our population to those smokers with complete data on socioeconomic status (SES) variables (i.e., car access, education, house tenure, household equivalized income, and neighborhood deprivation) and run the same analysis as in Model 1 (i.e., not yet adjusted for SES). Thus, the difference in RR compared with Model 1 (from 5.38 to 6.23, or 16 %) is due to selection bias from exclusion of those smokers with missing data

In Model 3, we use the same population as for Model 2, but now adjust for SES. Thus, the change in RR compared with Model 2 (from 6.23 to 5.83, or a 6 % reduction) represents the effect of confounding by SES

To estimate the best RR to this point, we assume the 6 % reduction due to confounding by SES would have applied to the Model 1 RR of 5.38 (i.e., before selection bias was introduced); this gives a best RR to this point of 5.03. We assume the standard error of $\ln[RR]$ from Model 3 applies to this RR of 5.03, giving a 95 % confidence interval of 3.88–6.53

Next, we adjust the RR further for misclassification bias of the smoking exposure. We use the observed association between RRs with and without adjustment of misclassification bias (as shown in Fig. 1); specifically, we use the regression equation for $\ln[\text{smoking misclassification-adjusted RR}]$ (y axis) on $\ln[\text{crude RR}]$ (x axis). Thus, our RR with adjustment for misclassification of smoking exposure (shown in the second to last, or “exposure misclassification” column, of Table 3) is:

$$\text{Exp}(1.1075 \times \ln[5.03] + 0.0264) = 6.14.$$

In calculating confidence intervals around this RR, we need to take account of uncertainty in the bias parameters themselves.

Uncertainty in misclassification of smoking exposure is captured in the Dirichlet distribution for the misclassification matrix (see Web Appendix 1). Across all our quantitative bias analyses (QBA) for exposure misclassification in Web Appendix 1, we found that the total uncertainty interval was on average 8 % greater than the random error-only 95 % confidence interval. Accordingly, we inflate the confidence intervals around the previous RR estimate by 8 % to give a ‘total uncertainty interval’ of 4.55–8.30

The final step is to quantify any ‘net’ residual confounding by alcohol and obesity, based on a simple averaging of the impact of these confounding QBA (as explained in Web Appendix 1). These further suggested corrections are shown as possible ranges of absolute adjustment to the RR in the final column of Table 3. In the case of larynx, ear, and nasosinus cancer, our assessment was that no additional adjustment was necessary

The RR estimates for cancer of the larynx, ear, and nasosinus cancer showed a fairly substantial degree of selection bias and confounding. For other cancers, the above adjustments were applied only if the shift in RRs between Models 1 and 2, and Models 2 and 3, was greater than 3 % (see “Methods” for further details)

log–log scale (Fig. 1). We used the equation for this slope and the finding that the uncertainty interval on the log scale was 8 % wider than the random error confidence interval, to calculate exposure misclassification-adjusted relative

risks and uncertainty intervals (UI) for all cancers as shown in the second to last column of Table 3. The misclassification exposure-adjusted RRs were >10 % higher than the original estimates for bladder, esophagus, and oropharynx cancer and >20 % higher for lung and larynx/ear/nasosinus cancer.

Finally, we also explored likely residual confounding by alcohol and smoking for the 1996–2001 cohort for selected cancers (details in Web Appendix 1). The impact of residual confounding was usually negligible, with overestimation of associations due to obesity (less common among smokers) often partly offset by underestimation due to confounding by alcohol (higher among smokers). Net effects of this confounding are depicted in the final column of Table 3, suggesting that our best estimate exposure misclassification-adjusted RRs still underestimate the RR for endometrial, kidney, and esophageal cancer by up to 0.20. Conversely, we likely overestimate oropharynx association by as much as 0.4. But with the exception of endometrial cancer, these cancers still have moderate to strong RR associations with smoking.

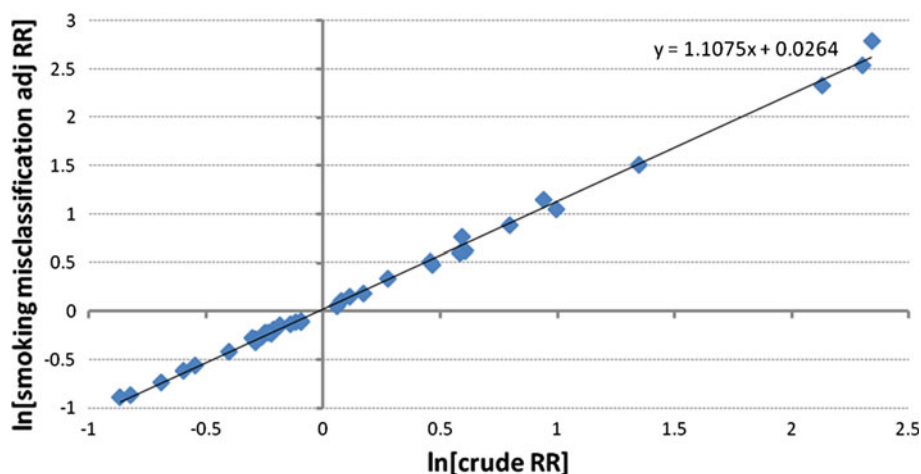
Substantively, the following patterns are evident in Table 3. Lung and larynx/ear/nasosinus cancers had markedly increased (RR > 6) incidence among current smokers. Bladder and esophagus have a clearly greater than doubling of incidence, and five cancers have a 20 % to twofold increased rate among current smokers with uncertainty intervals clearly excluding the null: cervix, kidney, oropharynx, pancreas, and stomach. Conversely, four cancers have at least a 20 % lower rate among current smokers with uncertainty intervals excluding the null: endometrial (empirical best estimates of 0.67; 95 % UI 0.56–0.79, and for 50 + years 0.60; 0.50–0.73; but due to residual confounding by obesity, the true RRs might be about 0.08 higher), CML (0.66; 0.44–0.99), melanoma (0.72; 0.65–0.81), and thyroid cancer (0.76; 0.58–1.00).

Eleven cancers had best RR estimates within 15 % of the null and with uncertainty intervals including the null: brain, breast, rectum, gallbladder and bile duct, Hodgkin’s, AML, CLL (and non-AML combined), myeloma, NHL (and follicular lymphoma), ovary, and testicular.

Colon cancer had a best empirical estimate RR of 0.89, with the uncertainty interval just excluding the null (0.81–0.98). However, our confounding bias analyses suggested that we were still slightly underestimating the RR, possibly to the extent that a corrected uncertainty interval would include the null.

The liver RRs varied markedly by ethnic grouping, with a strong 1.75 (1.37–2.24) association among European/other (probably slightly overestimated due to residual confounding by alcohol) but a null association among Māori/Pacific/Asian combined (0.80; 0.53–1.20).

Fig. 1 Plot of log of smoking misclassification-adjusted risk ratio by log of crude risk ratio, for 36 sex by age (45–64 and 65–84 year olds; European/other only) by cancer strata from the 1996–2001 cohort. Lung cancer among 65–84-year-old males excluded as outlier. See Web Appendix 1 for further details



Model 3 regressions were rerun to test for any significant interactions by sex and by cohort (Web Table 3). By sex, there was evidence of a stronger protective association of smoking for melanoma among males (RR 0.62; 0.56–0.69) compared to females (RR 0.84; 0.76–0.93). This stronger protective effect for males was evident in both the 1981–1986 and 1996–2001 cohorts (SRRs in Table 2). There was some suggestion that if there is any protective association for leukemia and its subtypes, it is only present in females. By cohort, there was a tendency for RRs to be further away from the null (be it less than or greater than 1.0) in the 1996–2001 cohort compared to the 1981–1986 cohort (Web Table 3).

Discussion

This study largely agrees with IARC or US-SG reports regarding associations of smoking with various cancers [1, 2]. Figure 2 plots the best empirical estimates from this study (i.e., second to last column of Table 3, plus additional adjustments for likely confounding by alcohol and obesity for kidney, liver (European/other only), esophageal, and oropharyngeal cancers) and recent meta-analyses (where they exist). The concordance is good. Considering category 1 and 2 cancers (positive causal association with smoking, Table 1), our estimates all have the lower 95 % confidence or uncertainty limits greater than 1.0. Our study also finds a strong protective association for endometrial cancer consistent with a previous meta-analysis [43]. We also find a protective association for thyroid cancer (RR 0.76; 95 % UI 0.58–1.00) adding weight to Mack et al.'s [33] pooled analysis that found an OR of 0.6 (0.6–0.7). One possible mechanism for both of these protective relationships is the anti-estrogenic effects of smoking [44].

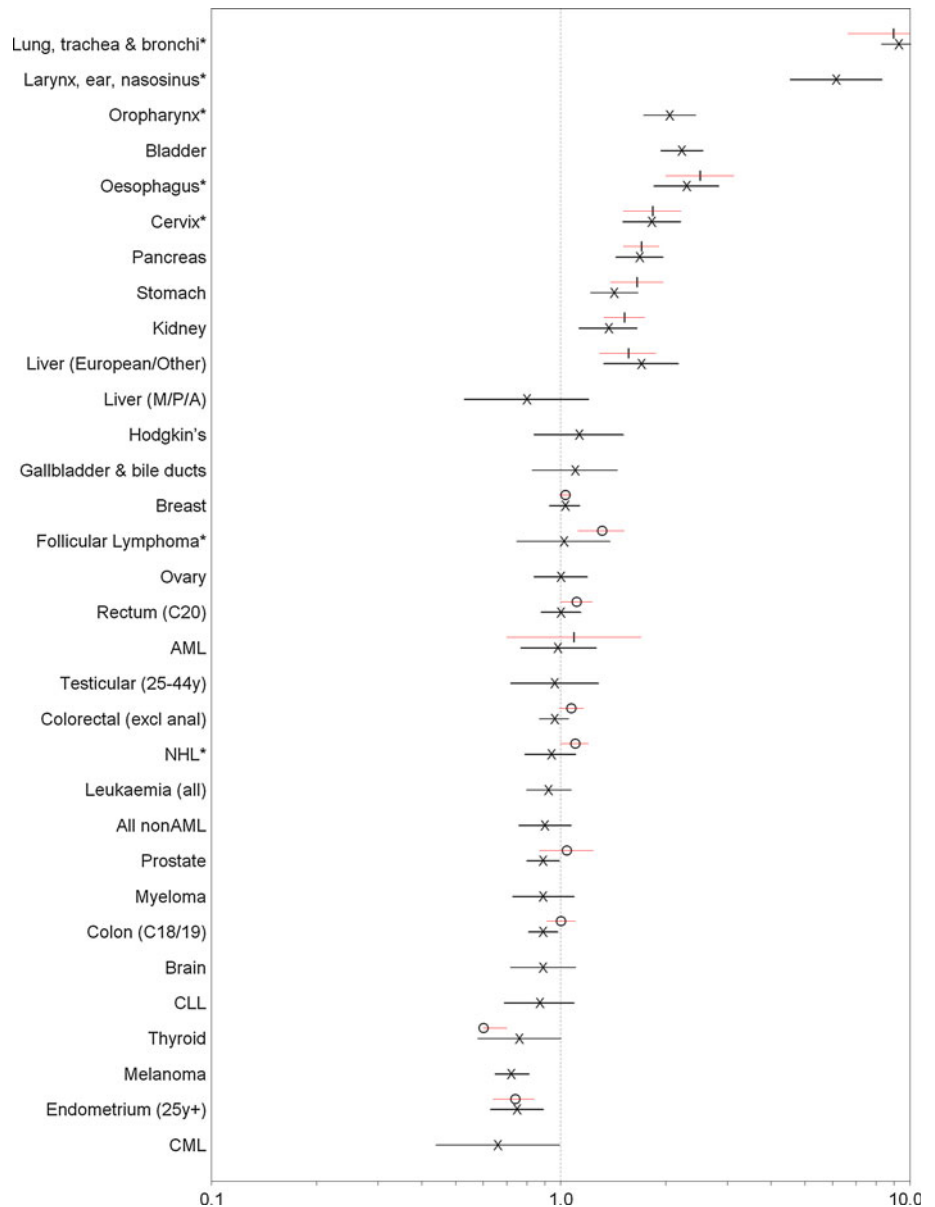
The category 4 “causality uncertain” cancers are best interpreted after a discussion of this study’s main strengths

and limitations: measurement error of the smoking exposure, and potential residual confounding.

The census smoking question that classified smoking status for our study is simple and does not capture intensity or duration. There were also likely to be some current smokers (inadvertently) reporting themselves as never smokers (or ex-smokers and hence excluded from our study), and vice versa some never smokers reporting themselves as current smokers. However, a key strength of our study is the inclusion of QBA which (assuming a valid bias model and parameters as detailed in Web Appendices 1 and 2) suggests negligible impact of exposure misclassification for most cancers and allows us to estimate exposure misclassification-adjusted estimates for all cancers (Table 3). While we have adjusted for bias from misclassification of the simple smoking question, further measurement error against a construct such as ‘pack-years’ was not possible; we assume that such correction would further strengthen the association for cancers already linked with smoking while having minimal effect on findings for other cancers. That said, we believe that misclassification bias is unlikely to be a major limitation for assessing ‘uncertain’ or marginal smoking–cancer associations in this study given the close agreement between our misclassification-corrected RR for cancer of the lung, trachea, and bronchus and that of Gandini et al.’s [25] meta-analysis of IARC data (Fig. 2), and the observation that misclassification of smoking status has little impact on RRs close to 1.0 (Fig. 1).

Our exploratory confounder bias adjustments provide considerable reassurance that residual confounding by alcohol and obesity is unlikely—except for our study underestimating the association of endometrial, kidney, and esophageal cancer with smoking and overestimating the association of liver (European/other only) and oropharyngeal cancer. But none of these four instances alter conclusions: Smoking will still be protective for endometrial cancer and increase risks for the other four.

Fig. 2 Forest plot of incidence rate ratios from: current CancerTrends study {best estimates from Table 3 [The best estimates from this study are those in the second to last column of Table 3 (i.e., misclassification bias adjusted), with the following four modification for likely residual confounding: addition of 0.08 to the endometrial cancer 25+ RR giving 0.75 (0.63–0.89); addition of 0.08 to the kidney cancer RR giving 1.37 (1.13–1.65); addition of 0.15 to the esophageal cancer RR giving 2.29 (1.85–2.83); subtraction of 0.25 from the oropharynx RR giving 2.05 (1.73–2.43); subtraction of 0.05 from European/other liver cancer RR giving 1.70 (1.33–2.17)]}; *solid black error bars*, with ‘multiplication sign’ central estimate; IARC data meta-analysis [25] (*red bars with vertical dash* central estimate); and miscellaneous meta-analyses (see text and footnotes to Table 1 for references) *red bars with ‘circle’* central estimate. *AML* acute myeloid leukemia, *CML* chronic myeloid leukemia, *CLL* chronic lymphocytic leukemia, *NHL* non-Hodgkin’s lymphoma. (Color figure online)



There was a tendency in our study for rate ratios to be further from the null in 1996–2001 than 1981–1986 (Table 2 and Web Table 3). This may be consistent with the 1996–2001 results reflecting a more mature smoking epidemic (especially for a long time-lag process such as cancer etiology) and less passive smoking risk in the never smokers, resulting in stronger rate ratios.

Returning to the substantive interpretation of our study findings for the category 4 “causality uncertain” cancers, one remaining source of bias deserves attention. The apparent negative association between smoking and prostate cancer may reflect more health conscious never-smokers being more likely to undergo prostate-specific antigen testing resulting in higher apparent incidence among this group, rather than a genuine protective effect. In contrast, it seems unlikely that

the observed protective association of smoking for melanoma is purely due to bias. A protective association for melanoma with RRs of 0.5–0.8 has been previously reported (although often with wide confidence levels including the null) [1, 31, 32, 45]. Curiously, the one other study to report results by sex also found a strong protective association among males, consistent with our study (Web Table 3). It is possible that the apparent protective effect is due to confounding by sunlight exposure, but this would require a strong association between low sunlight exposure and smoking within strata of socio-economic factors that we have already adjusted for. Such an association seems unlikely, and we know of no evidence for this in the NZ population. Thus, we would recommend re-categorizing melanoma as “Causality likely: protective association.” It is possible that because smoking has an

immunosuppressive effect [46], it protects melanocytes from UV radiation-induced inflammatory damage [32].

Smoking is recognized as a causal factor for myeloid leukemia by both IARC [1] and the US-SG [2], and the latter limits this to acute disease (AML). However, we did not find smoking to be significantly associated with AML, consistent with a recent meta-analysis for all myeloid leukemia (RR 1.09; CI 0.70–1.70) [47]. The apparently protective association for CML (best empirical estimate 0.66; 95 % UI 0.44–0.99) found in our study is inconsistent with the majority of international research, which suggest a null or possibly positive association [23, 24]. Ours may well have been a chance finding (consistent with an upper uncertainty limit of 0.99 and measurement of multiple associations).

Regarding liver cancer, our study discloses an important interaction of smoking with ethnicity (Table 3). It seems likely that among Māori/Pacific/Asian where hepatitis B is involved in the etiology of over 80 % of cases (compared to 6 % or less among European/other) [29], smoking is not a necessary causal cofactor with hepatitis B. However, our finding of an approximately 70 % increased risk among European/other smokers is consistent with the RR of 1.56 (95 % CI 1.29–1.87) from the meta-analysis of IARC data [47].

Finally, our study finds no compelling evidence of a link between smoking and breast, rectal, colon, gallbladder and bile duct, myeloma, brain, Hodgkin's, NHL, ovary and testicular cancers, and follicular lymphoma. Regarding colon, rectal, and breast cancers, there is some evidence that cancer risk increases with duration of smoking [10, 17, 19]. However, an increased risk with increased duration of smoking is also the case for other cancers that we do detect moderate or strong associations with, and most people initiate smoking by age 20. Our study cannot separately detect long-run from short-run causation, but it does seem unlikely that our study would produce null findings for colorectal and breast cancer—but not other cancers—if this hypothesis was true.

Conclusion

Our large study with measured and unmeasured confounder adjustment, and smoking misclassification adjustment, confirms the strong association of smoking with many cancers. However, our study also strengthens the evidence for protective associations of smoking for thyroid cancer and melanoma. We find no compelling association of smoking with breast and colorectal cancer. Contrary to recent statements, we find no association of smoking with AML. Our study also demonstrates how QBA can be applied to registry data and to strengthening causal inference in analyses on such data.

Acknowledgments Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the authors, not Statistics New Zealand.

Ethical standards The Wellington Ethics Committee granted ethics approval for CancerTrends (Ref 04/10/093).

Funding The authors received funds from Health Research Council of New Zealand.

Conflict of interest JJB is founder and owner of Epigear, which sells the Ersatz software used in the analysis.

References

1. International Agency for Research on Cancer (2004) Tobacco smoke and involuntary smoking. *IARC Monogr Eval Carcinog Risks Hum* 83:1–1187
2. Office of the Surgeon General (2004) The health consequences of smoking: a report of the Surgeon General. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Washington, DC
3. Blakely T et al (2010) CancerTrends: trends in cancer incidence by ethnic and socioeconomic group, New Zealand 1981–2004. University of Otago, and Ministry of Health, Wellington. www.wnmeds.ac.nz/academic/dph/research/HIRP/CancerTrends/CancerPublications.html
4. Blakely T, Wilson N (2005) The contribution of smoking to inequalities in mortality by education varies over time and by sex: two national cohort studies, 1981–84 and 1996–99. *Int J Epidemiol* 34(5):1054–1062
5. Hunt D et al (2005) The smoking-mortality association varies over time and by ethnicity in New Zealand. *Int J Epidemiol* 34:1020–1028
6. Blakely T et al (2006) What is the contribution of smoking and socioeconomic position to ethnic inequalities in mortality in New Zealand? *Lancet* 368(9529):44–52
7. Salmond C et al (2012) A decade of tobacco control efforts in New Zealand (1996–2006): impacts on inequalities in census-derived smoking prevalence. *Nicotine Tob Res* 14(6):664–673
8. Edwards R et al (2012) Setting a good example? Changes in smoking prevalence among key occupational groups in New Zealand: evidence from the 1981 and 2006 censuses. *Nicotine Tob Res* 14(3):329–337
9. Cornfield J et al (1959) Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst* 22:173–203
10. Lash T, Fox M, Fink A (2009) Applying quantitative bias analysis to epidemiological data. Springer, New York
11. Greenland S, Lash T (2008) Bias analysis. In: Rothman K, Greenland S, Lash T (eds) *Modern epidemiology*. Lippincott Williams & Wilkins, Philadelphia, pp 345–380
12. Greenland S (2009) Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. *Int J Epidemiol* 38:1662–1673
13. Collaborative Group on Hormonal Factors in Breast Cancer (2002) Alcohol, tobacco and breast cancer—collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *Br J Cancer* 87(11):1234–1245

14. Terry PD, Rohan TE (2002) Cigarette smoking and the risk of breast cancer in women: a review of the literature. *Cancer Epidemiol Biomarkers Prev* 11(10 Pt 1):953–971
15. Reynolds P et al (2004) Active smoking, household passive smoking, and breast cancer: evidence from the California Teachers Study. *J Natl Cancer Inst* 96(1):29–37
16. Cui Y, Miller AB, Rohan TE (2006) Cigarette smoking and breast cancer risk: update of a prospective cohort study. *Breast Cancer Res Treat* 100(3):293–299
17. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans (2012) A review of human carcinogens: part E—personal habits and indoor combustions. International Agency for Research on Cancer, Lyon
18. Botteri E et al (2008) Smoking and colorectal cancer: a meta-analysis. *JAMA* 300(23):2765–2778
19. Liang PS, Chen TY, Giovannucci E (2009) Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. *Int J Cancer* 124(10):2406–2415
20. Peach HG, Barnett NE (2001) Critical review of epidemiological studies of the association between smoking and non-Hodgkin's lymphoma. *Hematol Oncol* 19(2):67–80
21. Morton LM et al (2005) Cigarette smoking and risk of non-Hodgkin lymphoma: a pooled analysis from the International Lymphoma Epidemiology Consortium (interlymph). *Cancer Epidemiol Biomarkers Prev* 14(4):925–933
22. Adami J et al (1998) Smoking and the risk of leukemia, lymphoma, and multiple myeloma (Sweden). *Cancer Causes Control* 9(1):49–56
23. Brownson RC, Novotny TE, Perry MC (1993) Cigarette smoking and adult leukemia. A meta-analysis. *Arch Intern Med* 153(4):469–475
24. Siegel M (1993) Smoking and leukemia: evaluation of a causal hypothesis. *Am J Epidemiol* 138(1):1–9
25. Gandini S et al (2008) Tobacco smoking and cancer: a meta-analysis. *Int J Cancer* 122(1):155–164
26. Kurian AW et al (2005) Histologic types of epithelial ovarian cancer: have they different risk factors? *Gynecol Oncol* 96(2):520–530
27. Jordan SJ et al (2006) Does smoking increase risk of ovarian cancer? A systematic review. *Gynecol Oncol* 103(3):1122–1129
28. Collaborative Group on Epidemiological Studies of Ovarian Cancer (2012) Ovarian cancer and smoking: individual participant meta-analysis including 28 114 women with ovarian cancer from 51 epidemiological studies. *Lancet Oncol* 13(9):946–956
29. Blakely T et al (1999) Hepatitis B carriage explains the excess rate of hepatocellular carcinoma for Maori, Pacific Island, and Asian people compared to Europeans in New Zealand. *Int J Epidemiol* 28:204–210
30. Lee YC et al (2009) Meta-analysis of epidemiologic studies on cigarette smoking and liver cancer. *Int J Epidemiol* 38(6):1497–1511
31. Freedman DM et al (2003) Risk of melanoma in relation to smoking, alcohol intake, and other factors in a large occupational cohort. *Cancer Causes Control* 14(9):847–857
32. Odenbro A et al (2007) The risk for cutaneous malignant melanoma, melanoma in situ and intraocular malignant melanoma in relation to tobacco use and body mass index. *Br J Dermatol* 156(1):99–105
33. Mack WJ et al (2003) A pooled analysis of case-control studies of thyroid cancer: cigarette smoking and consumption of alcohol, coffee, and tea. *Cancer Causes Control* 14(8):773–785
34. Atkinson J et al (2010) CancerTrends technical report 1. Linkage of Census and Cancer Registrations, 1981–2004. University of Otago, Wellington
35. Salmond C, Crampton P (2001) NZDep96: what does it measure? *Soc Policy J N Z* 17:82–100
36. Heller W et al (1998) Misclassification of smoking in a follow-up population study in southern Germany. *J Clin Epidemiol* 51(3):211–218
37. Ministry of Health (2004) A portrait of health: key results of the 2002/03 New Zealand Health Survey. Ministry of Health, Wellington
38. Renehan AG et al (2008) Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 371(9612):569–578
39. Key J et al (2006) Meta-analysis of studies of alcohol and breast cancer with consideration of the methodological issues. *Cancer Causes Control* 17(6):759–770
40. Fedirko V et al (2011) Alcohol drinking and colorectal cancer risk: an overall and dose–response meta-analysis of published studies. *Ann Oncol* 22(9):1958–1972
41. Bagnardi V et al (2001) Alcohol consumption and the risk of cancer: a meta-analysis. *Alcohol Res Health* 25(4):263–270
42. Tramacere I, La Vecchia C, Negri E (2011) Tobacco smoking and esophageal and gastric cardia adenocarcinoma: a meta-analysis. *Epidemiology* 22(3):344–349
43. Zhou B et al (2008) Cigarette smoking and the risk of endometrial cancer: a meta-analysis. *Am J Med* 121(6):501.e3–508.e3
44. Baron JA, La Vecchia C, Levi F (1990) The antiestrogenic effect of cigarette smoking in women. *Am J Obstet Gynecol* 162(2):502–514
45. Westerdahl J et al (1996) Risk of malignant melanoma in relation to drug intake, alcohol, smoking and hormonal factors. *Br J Cancer* 73(9):1126–1131
46. Sopori M (2002) Effects of cigarette smoke on the immune system. *Nat Rev Immunol* 2(5):372–377
47. Gandini S et al (2008) Tobacco smoking and cancer: a meta-analysis. *Int J Cancer* 122(1):155–164
48. Appleby P et al (2006) Carcinoma of the cervix and tobacco smoking: collaborative reanalysis of individual data on 13,541 women with carcinoma of the cervix and 23,017 women without carcinoma of the cervix from 23 epidemiological studies. *Int J Cancer* 118(6):1481–1495
49. Vineis P et al (2004) Tobacco and cancer: recent epidemiological evidence. *J Natl Cancer Inst* 96(2):99–106
50. Hunt JD et al (2005) Renal cell carcinoma in relation to cigarette smoking: meta-analysis of 24 studies. *Int J Cancer* 114(1):101–108
51. Huncharek M et al (2010) Smoking as a risk factor for prostate cancer: a meta-analysis of 24 prospective cohort studies. *Am J Public Health* 100(4):693–701