BMJ Open

The burden of cancer attributable to modifiable risk factors: the Australian cancer-PAF cohort consortium

Maria E Arriaga,[1] Claire M Vajdic,[1] Karen Canfell,[2,3,4] Robert MacInnis,[5] Peter Hull,[1] Dianna J Magliano,[6] Emily Banks,[7] Graham G Giles,[5] Robert G Cumming,[3,8] Julie E Byles,[9] Anne W Taylor,[10] Jonathan E Shaw,[11] Kay Price,[12] Vasant Hirani,[3,13] Paul Mitchell,[14] Barbara-Ann Adelstein,[4] Maarit A Laaksonen[1]

To cite: Arriaga ME, Vajdic CM, Canfell K, *et al*. The burden of cancer attributable to modifiable risk factors: the Australian cancer-PAF cohort consortium. *BMJ Open* 2017;**7**:e016178. doi:10.1136/bmjopen-2017-016178

► Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2017-016178).

Received 30 January 2017
Revised 28 March 2017
Accepted 11 April 2017

For numbered affiliations see end of article.

Correspondence to
Dr. Maarit A Laaksonen;
m.laaksonen@unsw.edu.au

ABSTRACT

Purpose To estimate the Australian cancer burden attributable to lifestyle-related risk factors and their combinations using a novel population attributable fraction (PAF) method that accounts for competing risk of death, risk factor interdependence and statistical uncertainty.

Participants 365 173 adults from seven Australian cohort studies. We linked pooled harmonised individual participant cohort data with population-based cancer and death registries to estimate exposure-cancer and exposure-death associations. Current Australian exposure prevalence was estimated from representative external sources. To illustrate the utility of the new PAF method, we calculated fractions of cancers causally related to body fatness or both tobacco and alcohol consumption avoidable in the next 10 years by risk factor modifications, comparing them with fractions produced by traditional PAF methods.

Findings to date Over 10 years of follow-up, we observed 27 483 incident cancers and 22 078 deaths. Of cancers related to body fatness (n=9258), 13% (95% CI 11% to 16%) could be avoided if those currently overweight or obese had body mass index of 18.5–24.9 kg/m^2. Of cancers causally related to both tobacco and alcohol (n=4283), current or former smoking explains 13% (11% to 16%) and consuming more than two alcoholic drinks per day explains 6% (5% to 8%). The two factors combined explain 16% (13% to 19%): 26% (21% to 30%) in men and 8% (4% to 11%) in women. Corresponding estimates using the traditional PAF method were 20%, 31% and 10%. Our PAF estimates translate to 74 000 avoidable body fatness-related cancers and 40 000 avoidable tobacco- and alcohol-related cancers in Australia over the next 10 years (2017–2026). Traditional PAF methods not accounting for competing risk of death and interdependence of risk factors may overestimate PAFs and avoidable cancers.

Future plans We will rank the most important causal factors and their combinations for a spectrum of cancers and inform cancer control activities.

Strengths and limitations of this study

- A large, population-based, pooled prospective cohort with broad demographic and geographical coverage and individual participant data.
- Risk factor exposure prevalence estimates obtained from representative contemporary data sources to enhance the accuracy of population attributable fraction (PAF) estimates.
- The first cancer-PAF estimates from large-scale cohort study data that account for competing risk of death.
- Cancer-PAF estimates for the simultaneous effects of multiple risk factors, thereby accounting for their interdependence.
- CIs computed to show uncertainty in PAF estimates and differences between population subgroups.
- Estimates of the future numbers of cancers in Australia preventable by adherence to current recommendations for a healthy lifestyle.
- Evidence to underpin future evaluations of potential public health policies and interventions designed to reduce the cancer burden.
- Further improvements in the PAF methods are needed to incorporate the time the risk factor modification takes to be realised and the uncertainty in the exposure prevalence estimates.
- Larger cohort populations are needed to provide reliable data on some of the rarer cancers, cancer subtypes, risk factor combinations and population subgroups.

INTRODUCTION

Cancer is the leading cause of disease burden and death in Australia.[1 2] One of the principal strategies for reducing this burden is to target the key preventable causal factors, focusing activities where the association is strong, the exposure is common, and by considering the combination of both these factors overall and in population subgroups. The disease burden measure population attributable fraction (PAF) can be used to estimate the proportion of cancers that could be prevented if exposure to its risk factors were removed or reduced.[3 4] PAF accounts for both the strength of the exposure-cancer association and the exposure prevalence in

the population of interest. PAFs are increasingly used to evaluate the national, regional and global burden of cancer and to advocate for changes in public health policy and activity settings to reduce the prevalence of causal risk factors.[5] However, limitations in both the available data and the methods used restrict the accuracy of the PAF estimates and the scope of the conclusions.

Most prior cancer-PAF studies have relied on published exposure-cancer associations. As risk factor interaction and population subgroup analyses are rarely available, overall PAF estimates for individual risk factors dominate the literature.[6] Even where estimates are available, differences in the measurement and categorisation of risk factors and modelling approaches may limit their comparability. Most studies that have estimated PAFs for combined effects of risk factors have assumed independence between carcinogenic exposures.[5 7 8] Yet, in reality, modifiable lifestyle-related risk factors can interact to cause cancer and their effect may be higher for certain subgroups.[9] Moreover, these risk factors tend to co-occur or cluster, further adding to the burden of both cancer and death, and the effect of modifying one risk factor may be mediated by changes in other risk factors.[9–11]

PAFs are best estimated from cohort studies in which the risk factor exposure measurement precedes the cancer incidence.[12] Cohort studies also allow ascertainment of multiple outcomes related to an exposure and thus permit analyses to account for potential competing risks, such as death, which can alter PAF results.[13] To our knowledge, no previous cancer-PAF study has accounted for competing risk of death. This can be critical for cancers where established risk factors also predict risk of death from other causes and risk factor modifications will thus affect both outcomes. In addition, most previous cancer-PAF cohort studies have estimated exposure prevalence from the cohort population even when it has not been sampled to be representative of the target population of interest.[14] This hinders both generalisation and comparison of the findings. Finally, CIs for PAF estimates are often not provided, precluding an evaluation of their precision and also differences between population subgroups.

We addressed these deficiencies by applying our method[13 15] for estimating PAF and its CI for cancer incidence, allowing analysis of the simultaneous effects of multiple factors and accounting for competing risk of death, to an Australian cohort consortium and representative external exposure prevalence data.

COHORT CONSORTIUM DESCRIPTION

Australian cancer-PAF cohort consortium

The eligibility criteria for inclusion in the consortium were: well-established population-based Australian prospective cohort studies with comprehensive information on modifiable lifestyle-related exposures at baseline. Seven cohort studies met these criteria: Melbourne Collaborative Cohort Study (MCCS),[16] Blue Mountains Eye Study (BMES),[17] Australian Longitudinal Study on Women's Health (ALSWH),[18] Australian Diabetes, Obesity and Lifestyle Study (AusDiab),[19] North West Adelaide Health Study (NWAHS),[20] Concord Health and Ageing in Men Project (CHAMP)[21] and 45 and Up Study (45&Up).[22] Together they formed a study sample of 369 515 adult Australians of different ages covering the adult lifespan (table 1). Pooling of the cohorts identified 2457 people enrolled in more than one cohort, leaving a final population of 367 058 individuals, 365 173 with consent for record linkage.

The cohorts recruited participants between 1990 and 2009 (table 1). Only one cohort, AusDiab with recruitment from 1999, was designed to include a sample representative of the Australian population. Therefore, we used the latest representative external data sources to obtain contemporary age- and sex-specific risk factor prevalence estimates. These sources included the National Health Surveys (NHS) conducted 2014–2015 (NHS3),[23] 2004-2005 (NHS2)[24] and 2001 (NHS1),[25] the National Drug Strategy Household Survey (NDSHS) conducted in 2013[26] and the Learning how Australians Deal with menopausal sYmptoms (LADY) Survey conducted in 2013[27] (tables 1, 2 and 3), for which de-identified unit record data were available to generate the required exposure prevalences.

Data collection and harmonisation

All cohort studies collected baseline information on demographic, medical, lifestyle-related and hormonal exposures through self-completed questionnaires and some also through interviews and medical examinations (MCCS, BMES, AusDiab, NWAHS and CHAMP). We harmonised all available information on the relevant exposures across the cohort studies and the external data sources to the greatest extent possible (tables 2 and 3).

The modifiable exposures examined were regular smoking, alcohol consumption, body fatness (BMI ≥25 kg/m^2), physical activity, fruit consumption, vegetable consumption, red and processed meat consumption, oral contraceptive (OC) use, menopausal hormone therapy (MHT) use and breastfeeding. We classified the lifestyle exposures to match current Australian recommendations for healthy living, that is, not smoking, drinking no more than two standard alcoholic drinks per day (ie, 20 g of alcohol per day), maintaining healthy weight (BMI 18.5–24.9 kg/m^2), doing at least 150 min of moderate or 75 min of vigorous physical activity per week, eating at least two serves (ie, 300 g) of fruits and five serves (ie, 375 g) of vegetables per day, and not eating more than two serves (130 g) of either red or processed meat 3–4 times a week (table 2).

We also harmonised non-modifiable exposures such as age, gender, height, country of birth, marital status, education, socioeconomic status, urban–rural status, health insurance, reproductive history and personal and family medical history to allow population subgroup analyses and assessment of potential confounding factors (table 3).

Table 1 Characteristics of the data sources used in the Australian cancer-PAF cohort consortium

	Cohort data								Prevalence data				
Characteristic	MCCS	BMES	ALSWH	AusDiab	NWAHS	CHAMP	45&Up	Pooled	NHS1	NHS2	NHS3	NDSHS	LADY
Baseline year(s)	1990–1994	1992–1993	1996	1999–2000	1999–2003	2005–2007	2006–2009	1990–2009	2001	2004–2005	2014–2015	2013	2013
N	41 514	3654	40 310*	11 247	4056	1705†	267 029	367 058‡	17 918	19 501	14 560	22 696	4428
State/territory	VIC	NSW	All	All	SA	NSW	NSW	All	All	All	All	All	All
Age at baseline, mean (range)	55 (27–76)	66 (45–97)	45§ (18–75)	52 (25–95)	50 (18–90)	77 (70–97)	62 (45–≥100)	59 (18–≥100)	45 (18–80)	46 (18–85)	46 (18–85)	46 (18–84)	61 (50–71)
Women (%)	59	57	100	55	51	0	54	59	51	51	51	51	100

*1823 women did not consent for record linkage, leaving 38 487 women for the PAF analysis.
†66 men did not consent for record linkage, leaving 1639 men for the PAF analysis.
‡2457 individuals participated in more than one cohort study and were only included in the first cohort study they participated in, leaving 367 058 individuals in the pooled data. Of these, 1885 did not consent for record linkage, leaving 365 173 individuals for the PAF analysis.
§The ALSWH recruited three cohorts aged 18–23, 45–50 and 70–75 so the age distribution is not continuous.
45&Up, 45 and Up Study; ALSWH, Australian Longitudinal Study on Women's Health; AusDiab, Australian Diabetes, Obesity and Lifestyle Study; BMES, Blue Mountains Eye Study; CHAMP, Concord Health and Ageing in Men Project; LADY, Learning how Australians Deal with menopausal sYmptoms Survey; MCCS, Melbourne Collaborative Cohort Study; NDSHS, National Drug Strategy Household Survey; NHS, National Health Survey; NSW, New South Wales; NWAHS, North West Adelaide Health Study; PAF, population attributable fraction; SA, South Australia; VIC, Victoria.

Data linkage

We linked the pooled cohort to the population-based Australian Cancer Database (ACD) and National Death Index (NDI) to identify cancers and deaths. The ACD is a database of all primary, malignant cancers, except keratinocyte cancers, notified to State and Territory Cancer Registries in Australia since 1982. The NDI records all deaths registered in Australia since 1980. Both the ACD and NDI are maintained by the Australian Institute of Health and Welfare, which facilitates research by conducting record linkage using an established probabilistic linkage algorithm.[28]

In October 2016, the ACD and NDI records were available until the end of 2012, providing 8–22 years follow-up depending on the individual cohort (table 1).

Data analysis and statistical methods

We classified cancers on the basis of the International Classification of Diseases for Oncology codes. Only invasive cancers identified by data linkage were included, and people with a cancer registration prior to baseline were excluded from the analysis for that malignancy.

We defined follow-up as the time from baseline to the date of diagnosis of the cancer of interest, death or end of follow-up, whichever occurred first. The survival times were assumed to follow a parametric proportional hazards model with piecewise constant baseline hazard function.[29] Maximum likelihood estimation with iterative methods was used to obtain the parameter estimates and their estimated covariance matrices.[13] We expressed the strength of exposure-cancer and exposure-death associations adjusted for baseline age, sex and study as HRs and their 95% CIs. We computed the corresponding age- and sex-specific exposure prevalence estimates from the most contemporary representative external data source. Participants with missing data for the variables included in the model were excluded from the analyses. We then combined the maximum likelihood estimates and the exposure prevalence estimates to calculate the PAF point estimates using our recently developed PAF formula[13] accounting for competing risk of death. The asymptotic variance estimate of PAF was obtained using the delta method, and two-sided 95% CIs for the PAFs were calculated by applying a symmetrising complementary logarithmic transformation of PAF.[13] Our PAF method[13] and program[15] allows a flexible choice of the reference level for the hypothetical risk factor modification and simultaneous analysis of the effects of multiple risk factors. Both individual PAFs for modification of single risk factors and joint PAFs for modification of several risk factors can be calculated.

To illustrate the novel cancer-PAF estimation, we estimated the fractions of cancers causally related to (1) body fatness and (2) both tobacco and alcohol consumption attributable to these risk factors in Australia over the next 10 years. We restricted these analyses to the first 10-year follow-up to generate comparable estimates across the cohorts. We included only those cancers judged by

Table 2 List of main harmonised modifiable baseline risk factors for cohort studies and external prevalence data sources

	Cohort data						Prevalence data					
Risk factors	**MCCS**	**BMES**	**ALSWH**	**AusDiab**	**NWAHS**	**CHAMP**	**45&Up**	**NHS1**	**NHS2**	**NHS3**	**NDSHS**	**LADY**
Smoking												
Regular smoking (never, former, current)	√	√	√	√	√	√	√	√	√	√	√	√
Cigarettes/day	√	√	√	√	√	√	√	-	-	-	√	√
Time since quitting (years)	√	√	√	√	√	√	√	√	√	√	√	√
Duration (years)	√	√	√	√	√	√	√	√	√	√	√	√
Australian recommendation: no smoking	√	√	√	√	√	√	√	√	√	√	√	√
Alcohol consumption												
Drinks/day	√	√	√	√	√	√	√	√	√	√	√	√
Australian recommendation: ≤2 drinks/day	√	√	√	√	√	√	√	√	√	√	√	√
Body fatness												
BMI (kg/m^2)*	√	√	√	√	√	√	√	√	√	√	-	√
Australian recommendation: 18.5–24.9 kg/m^2	√	√	√	√	√	√	√	√	√	√	-	√
Physical activity												
Moderate activity (min/week)	~	√	~	√	√	~	√	√	√	√	-	~
Vigorous activity (min/week)	~	√	~	√	√	~	√	√	√	√	-	~
Australian recommendation: ≥150 min/week†	~	√	~	√	√	~	√	√	√	√	-	~
Diet												
Fruit consumption (serves/day)	√	√	-‡	√	-‡	-	√	√	√	√	-	-
Australian recommendation: ≥2 serves/day	√	√	-‡	√	-‡	-	√	√	√	√	-	-
Vegetable consumption (serves/day)	√	√	-‡	√	-‡	-	√	√	√	√	-	-
Australian recommendation: ≥5 serves/day	√	√	-‡	√	-‡	-	√	√	√	√	-	-
Red meat consumption (times/week)	√	-	-‡	~	-	-	√	-	-	-	-	-
Processed meat consumption (times/week)	√	-	-‡	~	-‡	-	√	-	-	-	-	-
Australian recommendation: ≤3–4 times/week	√	-	-‡	~	-‡	-	√	-	-	-	-	-
OC use												
OC use (never, former, current)	√	~	√	√	-	N/A	√	√	-	-	-	√
Duration (<5 years, ≥5 years)	√	-	√	√	-	N/A	√	~	-	-	-	-
MHT use												
MHT use (never, former, current)	√	~	√	√	-	N/A	√	√	√	-	-	√
Duration (<5 years, ≥5 years)	√	-	√	-	-	N/A	√	-	-	-	-	√
Breastfeeding												

Continued

Table 2 Continued

Risk factors	Cohort data						Prevalence data					
	MCCS	BMES	ALSWH	AusDiab	NWAHS	CHAMP	45&Up	NHS1	NHS2	NHS3	NDSHS	LADY
Duration (months)	√	-	-‡	-	-	N/A	√	√	-	-	-	√

√, available; ~, not comparable; -, not available; N/A, not applicable.
*Measured weight and height (MCCS, AusDiab, NWAHS, CHAMP and NHS3); self-reported weight and height (BMES, ALSWH, 45&Up, NHS1, NHS2 and LADY).
†≥ 150 min/week of moderate activity or ≥75min/week of vigorous activity or combination of the two.
‡Not available at baseline but available at later measurements.
45&Up, 45 and Up Study; ALSWH, Australian Longitudinal Study on Women's Health; AusDiab, Australian Diabetes, Obesity and Lifestyle Study; BMES, Blue Mountains Eye Study; BMI,body mass index; CHAMP, Concord Health and Ageing in Men Project; LADY, Learning how Australians Deal with menopausal sYmptoms Survey; MCCS, Melbourne Collaborative Cohort Study; MHT, menopausal hormone therapy; NDSHS, National Drug Strategy Household Survey; NHS, National Health Survey; NWAHS, North West Adelaide Health Study; OC, oral contraceptive.

the International Agency for Research on Cancer to be causally associated with body fatness (oesophageal adenocarcinoma, stomach, colorectal, liver, gallbladder, pancreas, postmenopausal breast, corpus uteri, ovary, renal-cell carcinoma, meningioma, thyroid and multiple myeloma) or with *both* tobacco and alcohol consumption (tongue, mouth, oropharynx, hypopharynx, other pharynx (excluding nasopharynx), oesophagus, colorectal, liver and larynx).[30 31] It should be noted that cancers of the lung and breast are not included here as they are not causally related to *both* tobacco and alcohol consumption. We estimated the individual contribution of body fatness, tobacco and alcohol consumption and the combined contribution of tobacco and alcohol consumption on the burden of the respective cancers. For body fatness, we evaluated scenarios in which (1) those currently obese or overweight had healthy weight and (2) those currently obese were overweight. For smoking, we evaluated scenarios in which (1) current and former smokers had never smoked and (2) current smokers were to quit and become former smokers. For alcohol consumption, we evaluated the scenario in which no-one drank more than two alcoholic drinks per day. We also evaluated potential effect modification of the contribution of smoking by alcohol consumption and all three risk factors by sex. We estimated the numbers of these cancers that could be avoided in Australia under these scenarios by multiplying the PAF estimates by the projected numbers of cancers over the next 10 years (2017–2026).[32 33]

To demonstrate the potential impact of our methodology on PAF estimates, we compared our results with PAF estimates produced by traditional methods,[3 4] adapted to cohort studies with survival data by replacing the relative risks (RRs) in the original formulas by HRs from survival models[34] that do not account for competing risk of death and that take a sequential approach to estimate the combined effect of multiple risk factors, assuming their independence.[7 14]

We carried out all statistical analyses using SAS V.9.4 and a publicly available PAF program based on SAS macros.[15]

FINDINGS TO DATE

Harmonisation and prevalence of lifestyle-related risk factors

Smoking, alcohol consumption and body fatness could be harmonised for all cohorts, while physical activity, fruit and vegetable consumption, red and processed meat consumption, OC and MHT use and breastfeeding were either not collected at baseline or could not be harmonised for some cohorts (table 2).

The participant age and sex distribution varied across the cohorts (table 1) as did the crude risk factor exposure prevalences (table 4), even in cohorts recruited around the same time. The sex- and age-stratified exposure prevalence estimates were more comparable but

Table 3 List of main harmonised non-modifiable baseline risk factors for cohort studies and external prevalence data sources

	Cohort data							Prevalence data				
Rish factors	**MCCS**	**BMES**	**ALSWH**	**AusDiab**	**NWAHS**	**CHAMP**	**45&Up**	**NHS1**	**NHS2**	**NHS3**	**NDSHS**	**LADY**
Age	√	√	√	√	√	√	√	√	√	√	√	√
Gender	√	√	√	√	√	√	√	√	√	√	√	√
Height (cm)*	√	√	√	√	√	√	√	√	√	√	-	√
Country of birth (Australia, overseas)	√	-	√	√	√	√	√	√	√	√	√	√
Marital status	√	√	√	√	√	√	√	√	√	√	√	√
Education (basic, intermediate, high)	√	√	√	√	√	√	√	√	√	√	√	√
Socioeconomic status (SEIFA)	√	√	√	√	√	√	√	√	√	√	√	√
Urban–rural status (ARIA)	√	√	√	√	√	√	√	√	√	√	√	√
Health insurance	-	-	√	√	-	-	√	√	√	√	-	-
Reproductive history												
Age at menarche	√	√	-§	-	-	N/A	√	-	-	-	-	√
Age at first delivery	√	-	-§	√	-	N/A	√	-	-	-	-	√
Parity	√	√	√	√	-	N/A	√	√	√	-	-	√
Menopausal status	√	√	√	√	-§	N/A	√	-	-	-	-	√
Age at menopause	√	√	-§	√	-§	N/A	√	-	-	-	-	√
Family history of cancer†	√	-	-§	-	-	√	√	-	-	-	-	√
Medical history												
Cancer screening‡	-	√	√	-	-	√	√	√	-	√	-	√
Diabetes	√	√	√	√	√	√	√	√	√	√	√	√

√, available; ~, not comparable; -, not available; N/A, not applicable.
*Measured height (MCCS, AusDiab, NWAHS, CHAMP and NHS3); self-reported height (BMES, ALSWH, 45&Up, NHS1, NHS2 and LADY).
†Any cancer (MCCS); prostate (CHAMP and 45&Up); breast and ovarian (45&Up and LADY), colorectal, lung, melanoma (45&Up).
‡Screening for breast, prostate, colorectal (45&Up and NHS3); breast only (BMES, ALSWH, NHS1 and LADY); prostate only (CHAMP).
§Not available at baseline but available at later measurements.
45&Up, 45 and Up Study; ALSWH, Australian Longitudinal Study on Women's Health; ARIA, Accessibility/Remoteness Index of Australia; AusDiab, Australian Diabetes, Obesity and Lifestyle Study; BMES, Blue Mountains Eye Study; CHAMP, Concord Health and Ageing in Men Project; LADY, Learning how Australians Deal with menopause sYmptoms Survey; MCCS, Melbourne Collaborative Cohort Study; NDSHS, National Drug Strategy Household Survey; NHS, National Health Survey; NWAHS, North West Adelaide Health Study; SEIFA, Socio-Economic Indexes for Areas.

Table 4 Crude prevalence (%) of main lifestyle-related risk factors for cancer at baseline for cohort studies and external prevalence data sources

	Men									Women								
	Cohort data						Prevalence data			Cohort data						Prevalence data		
	MCCS	BMES	AusDiab	NWAHS	CHAMP	45&Up	NHS1	NHS2	NHS3	MCCS	BMES	ALSWH	AusDiab	NWAHS	45&Up	NHS1	NHS2	NHS3
Risk factors	**1990–1994**	**1992–1993**	**1999–2001**	**1999–2003**	**2005–2007**	**2006–2009**	**2001**	**2004-2005**	**2014–2015**	**1990–1994**	**1992–1993**	**1996**	**1999–2001**	**1999–2003**	**2005–2008**	**2001**	**2004-2005**	**2014–2015**
Age (median)*	56	66	49	49	76	62	43	43	45	55	66	47	41	43	59	43	44	45
Smoking†																		
Never regular	41	32	49	43	37	49	44	46	51	69	61	62	64	53	65	58	59	63
Former regular	45	52	34	34	57	44	30	30	32	22	26	24	23	30	28	22	23	25
Current regular	15	15	17	23	6	8	25	24	17	9	13	14	12	17	7	20	18	12
Alcohol consumption																		
≤2 drinks/day	71	80	76	83	85	76	71	67	74	91	95	95	95	97	94	92	88	91
>2 drinks/day	29	20	24	17	15	24	29	33	26	9	5	5	5	3	6	8	12	9
Body fatness (BMI; kg/m^2)†,‡																		
<18.5	0	1	0	1	1	1	1	1	1	1	3	5	1	2	2	5	4	2
18.5–24.9	27	39	31	31	23	31	41	37	28	41	43	56	45	40	42	53	51	42
25.0–29.9	53	46	49	43	49	47	42	43	42	36	34	26	31	30	33	25	28	29
≥30.0	19	14	20	25	27	22	16	19	28	22	20	13	23	29	23	17	17	27
Physical activity																		
<150 min/week§	~	75	49	66	~	36	46	57	69	~	85	-	62	82	32	59	65	79
≥150 min/week§	~	25	51	34	~	64	54	43	31	~	15	-	38	18	68	41	35	21
Fruit consumption																		
<2 serves/day	20	64	60	-	-	49	54	52	56	12	60	-	48	-	35	42	40	45
≥2 serves/day	80	36	40	-	-	51	46	48	44	88	40	-	52	-	65	58	60	55
Vegetable consumption																		
<5 serves/day	53	41	92	-	-	76	73	88	92	37	37	-	84	-	61	66	84	90
≥5 serves/day	47	59	8	-	-	24	27	12	8	63	63	-	16	-	39	34	16	10

~, not comparable; -, not available.
*Prevalences by age groups are given in online supplementary table 1.
†Some percentages do not add up to 100 because of rounding.
‡Measured weight and height (MCCS, AusDiab, NWAHS, CHAMP and NHS3); self-reported weight and height (BMES, ALSWH, 45&Up, NHS1, NHS2 and LADY).
§≥ 150 min/week of moderate activity or ≥75 min/week of vigorous activity or combination of the two.
45&Up, 45 and Up Study; ALSWH, Australian Longitudinal Study on Women's Health; AusDiab, Australian Diabetes, Obesity and Lifestyle Study; BMES, Blue Mountains Eye Study; BMI, body mass index; CHAMP, Concord Health and Ageing in Men Project; LADY, Learning how Australians Deal with menopause sYmptoms Survey; MCCS, Melbourne Collaborative Cohort Study; NHS, National Health Survey; NWAHS, North West Adelaide Health Study.

Table 5 Exposure prevalence and HRs for cancers causally related to body fatness or both tobacco and alcohol over 10-year follow-up

	Men and women		Men		Women	
Risk factors	**Prevalence***	**HR (95% CI)†**	**Prevalence***	**HR (95% CI)†**	**Prevalence***	**HR (95% CI)†**
Body fatness (BMI; kg/m^2)‡,§						
Underweight (<18.5)	2%	0.84 (0.69 to 1.01)	1%	1.22 (0.80 to 1.87)	2%	0.78 (0.63 to 0.96)
Healthy weight (18.5–24.9)	35%	1	28%	1	42%	1
Overweight (25.0–29.9)	36%	1.14 (1.09 to 1.20)	42%	1.18 (1.08 to 1.29)	29%	1.13 (1.06 to 1.20)
Obese (≥30.0)	28%	1.36 (1.29 to 1.44)	28%	1.42 (1.28 to 1.58)	27%	1.35 (1.26 to 1.44)
Smoking¶						
Never regular	57%	1	51%	1	63%	1
Former regular	28%	1.29 (1.20 to 1.37)	32%	1.44 (1.31 to 1.57)	25%	1.20 (1.08 to 1.32)
Current regular	15%	1.59 (1.43 to 1.77)	17%	1.96 (1.70 to 2.26)	12%	1.29 (1.09 to 1.52)
Alcohol consumption¶						
≤2 drinks/day	83%	1	74%	1	91%	1
>2 drinks/day	17%	1.38 (1.27 to 1.50)	26%	1.41 (1.28 to 1.55)	9%	1.27 (1.06 to 1.51)
Smoking by alcohol consumption						
≤2 drinks/day						
Never regular smoker	51%	1	42%	1	59%	1
Former regular smoker	22%	1.24 (1.15 to 1.34)	22%	1.35 (1.22 to 1.50)	21%	1.19 (1.07 to 1.32)
Current regular smoker	10%	1.41 (1.24 to 1.61)	10%	1.74 (1.44 to 2.09)	10%	1.21 (1.00 to 1.45)
>2 drinks/day						
Never regular smoker	6%	1	9%	1	4%	1
Former regular smoker	7%	1.44 (1.21 to 1.72)	10%	1.53 (1.25 to 1.86)	4%	1.20 (0.81 to 1.80)
Current regular smoker	4%	2.09 (1.67 to 2.62)	7%	2.24 (1.73 to 2.89)	2%	1.73 (1.06 to 2.81)

*From the National Health Survey (NHS) 2014–2015.
†Adjusted for age, sex and study.
‡Some percentages do not add up to 100 because of rounding.
§Cancers causally related to body fatness are oesophageal adenocarcinoma (C15, histology codes 8140–8576), cancers of stomach (C16), colon (C18), rectum (C19-20), liver (C22), gallbladder (C23) and pancreas (C25), postmenopausal breast cancer (C50), cancers of corpus uteri (C54) and ovary (C56), renal-cell carcinoma (C64, histology codes 8050, 8140, 8260, 8270, 8280–8312, 8316–8320, 8340–8344), meningioma (C70), thyroid cancer (C73) and multiple myeloma (C90). After exclusion of individuals with prevalent body fatness-related cancer at baseline, 351 035 individuals were available for analyses and 328 438 (93.6%) of them had BMI information available.
¶Cancers causally related to both tobacco and alcohol are cancers of tongue (C01-02), mouth (C03-C06), oropharynx (C10), hypopharynx (C12-C13), other sites in pharynx (C14), oesophagus (C15), colon (C18), rectum (C19-C20), liver (C22) and larynx (C32). After exclusion of individuals with prevalent tobacco and alcohol-related cancer at baseline, 359 796 individuals were available for analyses and 356 117 (99.0%) of them had smoking, 351 119 (97.6%) alcohol consumption and 348 070 (96.7%) both smoking and alcohol consumption information available.
BMI, body mass index.

generally lower for the cohort studies than the representative external data sources from around the same time (see online supplementary table 1).

Exposure prevalence estimates from the representative external sources (NHS 2001, 2004 and 2014–2015; table 4) showed different temporal trends depending on specific risk factors. The overall prevalence (men and women) of current smoking decreased over time (22%, 21%, and 15%, respectively). The overall prevalence of consuming more than two alcoholic drinks a day (19%, 22% and 17%) and inadequate fruit consumption (48%, 46% and 50%) were relatively stable over time. The prevalence of body fatness (50%, 54% and 63%), physical inactivity (52%, 61% and 74%) and inadequate vegetable consumption (70%, 86% and 91%) increased over time. Currently, the prevalence of many modifiable lifestyle-related risk factors exceeds 50% of Australians and is generally higher in men than women (table 4).

The variable cohort age and sex distribution and the temporal trends in exposure prevalences demonstrate the need to use representative and most recent prevalence estimates for reliable PAF calculations. Red and processed meat consumption were the only exposures that could not be obtained from a representative external source (tables 2 and 3); we obtained these prevalence estimates from the largest and latest cohort (The 45 and Up Study) and will perform sensitivity analyses to assess the impact of the uncertainty in this measure.

Cancer and death cases in the pooled cohort

During the maximum 22-year follow-up of the pooled cohort (n=365 173; table 1) with mean age 59 years and 59% women, 35 860 incident cancers and 32 107 deaths were observed (see online supplementary table 2). The distribution of the cancers in the pooled cohort is similar to that for the Australian population.[2] During the first 10-year follow-up, we observed 27 483 cancers and 22 078 deaths. There were 9258 participants with a first primary cancer causally related to body fatness and 4283 participants with a first primary cancer causally related to both tobacco and alcohol (see online supplementary table 3). No significant heterogeneity between the cohort-specific HRs for cancers causally related to body fatness or both tobacco and alcohol was found (see online supplementary table 4).

Avoidable cancers causally related to body fatness, and both tobacco and alcohol

Individual and combined contributions of risk factors

According to our estimations, overweight and obesity (table 5) explain 13% (95% CI 11% to 16%) of the 10-year burden of cancers causally associated with body fatness (table 6). If those currently obese were overweight, 5% (95% CI 3% to 7%) of the burden could be avoided. For cancers causally related to both tobacco and alcohol consumption, 13% (95% CI 11% to 16%) is attributable to smoking and could be avoided if current and former smokers had never smoked (table 6). If current smokers were to quit, 3% (95% CI 1% to 4%) of the burden could be avoided. Drinking more than two alcoholic drinks per day explains 6% (95% CI 5% to 8%) of the burden. Excessive alcohol consumption combined with ever smoking explains 16% (95% CI 13% to 19%) and combined with current smoking 8% (95% CI 5% to 10%) of the burden of these cancers over the next 10 years. Current smokers who also consume more than two alcoholic drinks per day are at a significantly higher risk of these cancers compared with smokers whose alcohol consumption does not exceed two daily drinks (HR 2.09 vs 1.41; table 5) and would benefit much more from quitting smoking (PAF 9% (95% CI 4% to 15%) versus 1% (95% CI 0% to 3%)).

Contributions by sex

The contribution of body fatness to cancers causally related to this risk factor was 17% (95% CI 11% to 22%) for men and 12% (95% CI 9% to 15%) for women. The contribution of both smoking and excessive alcohol consumption to the burden of cancers causally related to both these exposures was even more pronounced for men (table 6). The PAFs for current and former smoking were 22% (95% CI 17% to 26%) for men versus 7% (95% CI 3% to 10%) for women and for consuming more than two alcoholic drinks per day 9% (95% CI 6% to 12%) versus 2% (95% CI 0% to 4%). Modifications to both of these risk factors could reduce the burden by 26% (95% CI 21% to 30%) for men and 8% (95% CI 4% to 11%) for women.

Comparison of novel and traditional PAF methods

Body fatness and alcohol consumption were weakly associated with death from causes other than the cancers of interest, whereas smoking was a moderate risk factor for mortality during the 10-year follow-up (HR 1.36 for former smokers and 2.23 for current smokers). Accordingly, the PAF estimates based on the novel and traditional methods differed most for smoking, especially for men (table 6). The differences between the two methods were larger when the combined effect of modifying both smoking and alcohol consumption was analysed. The point estimates given by the traditional method no longer fitted within the CIs produced by the novel method both for the overall population (20% vs 16% (95% CI 13% to 19%) and for men (31% vs 26% (95% CI 21% to 30%); however, the lack of CIs around the traditional estimates makes direct comparison difficult.

Projected avoidable numbers of cancers

Based on the projected Australian cancer incidence rates over the next 10 years, around 840 000 people will be diagnosed with cancer, of which over 570 000 are cancers causally related to body fatness and 250 000 cancers causally related to both tobacco and alcohol consumption. Of these, according to our PAF estimates, 74 000 cases can be attributed to body fatness, 32 000 to smoking, 15 000

Table 6 Fractions of cancers causally related to body fatness or both tobacco and alcohol attributable to the respective risk factors over 10 years of follow-up

	Men and women		Men		Women	
	PAF (95% CI)		PAF (95% CI)		PAF (95% CI)	
Behaviour modification	Competing risk method	Traditional method‡	Competing risk method	Traditional method‡	Competing risk method	Traditional method‡
Body fatness*						
Obese and overweight to healthy weight	13 (11 to 16)	13	17 (11 to 22)	17	12 (9 to 15)	12
Obese to overweight	5 (3 to 7)	5	5 (2 to 8)	6	5 (3 to 7)	5
Smoking†						
Current and former smokers to never smokers	13 (11 to 16)	15	22 (17 to 26)	23	7 (3 to 10)	8
Current smokers to former smokers	3 (1 to 4)	4	5 (3 to 8)	7	1 (-1, 2)	1
Alcohol consumption†						
>2 drinks/day to ≤2 drinks/day	6 (5 to 8)	6	9 (6 to 12)	10	2 (0 to 4)	2
Smoking and alcohol consumption						
Current and former smokers to never smokers and >2 drinks/day to ≤2 drinks/day	16 (13 to 19)	20	26 (21 to 30)	31	8 (4 to 11)	10
Current smokers to former smokers and >2 drinks/day to ≤2 drinks/day	8 (5 to 10)	10	12 (9 to 16)	16	2 (–1 to 5)	3
Smoking by alcohol consumption						
≤2 drinks/day						
Current and former smokers to never smokers	9 (6 to 12)	12	16 (11 to 20)	20	5 (2 to 9)	7
Current smokers to former smokers	1 (0 to 3)	2	3 (0 to 6)	5	0 (–2 to 2)	0
>2 drinks/day						
Current and former smokers to never smokers	29 (19 to 37)	36	32 (21 to 42)	41	17 (–7 to 35)	23

Continued

Table 6 Continued

Behaviour modification	Men and women		Men		Women	
	PAF (95% CI)		PAF (95% CI)		PAF (95% CI)	
	Competing risk method	Traditional method‡	Competing risk method	Traditional method‡	Competing risk method	Traditional method‡
Current smokers to former smokers	9 (4 to 15)	11	10 (4 to 16)	12	7 (−4 to 16)	8

*Cancers causally related to body fatness are oesophageal adenocarcinoma (C15, histology codes 8140–8576), cancers of stomach (C16), colon (C18), rectum (C19-20), liver (C22), gallbladder (C23) and pancreas (C25), postmenopausal breast cancer (C50), cancers of corpus uteri (C54) and ovary (C56), renal-cell carcinoma (C64, histology codes 8050, 8140, 8260, 8270, 8280–8312, 8316–8320, 8340–8344), meningioma (C70), thyroid cancer (C73) and multiple myeloma (C90).
†Cancers causally related to both tobacco and alcohol are cancers of tongue (C01-02), mouth (C03-C06), oropharynx (C10), hypopharynx (C12-C13), other sites in pharynx (C14), oesophagus (C15), colon (C18), rectum (C19-C20), liver (C22) and larynx (C32).
‡PAF for individual contribution of a risk factor is calculated by ∑Prevalence*(HR−1)/[1+∑Prevalence*(HR−1)] using information in table 5. The sum in the numerator is for the modified levels of exposure (eg, current and former smokers or just current smokers), the sum in the denominator is for all levels of exposure and the reference level with HR=1 the target level of modification (eg, never smokers or former smokers). PAF for the combined contribution of two risk factors is calculated by 1−(1−PAF1)*(1−PAF2).
PAF, population attributable fraction.

to consuming more than two alcoholic drinks per day and 40 000 to the latter two exposures combined. The number of cancers preventable through avoiding both smoking and excessive alcohol consumption are overestimated by 10 000 if the competing risk of death is not considered, and these two exposures are assumed to act independently.

STRENGTHS AND LIMITATIONS

Strengths

The large cohort and advanced PAF methodology enables analysis of both the individual and joint contribution of risk factors to the burden of cancer, both overall and in subgroups. In the next stage of this research, we will identify and rank the most harmful cancer risk factors and their combinations for specific cancers and evaluate the distribution of their burden. We will also evaluate the contribution of different risk factors across all cancers. We will use this epidemiological evidence to inform future health promotion and other cancer control activities.

Access to individual participant cohort data allowed us to harmonise risk factors, potential confounding factors and effect modifiers. This is expected to increase the comparability and accuracy of the PAF estimates. As recommended,[35] we documented our rigorous guidelines for data harmonisation to enable reproducibility and use in subsequent pooling efforts. We aligned our exposure classifications with the Australian recommendations for maintaining a healthy lifestyle, allowing consistency of risk communication.

Utilising corresponding risk factor exposure prevalence estimates from representative data sources also increased the accuracy of our PAF estimates. Cohort studies may have variable age and sex distributions, and they may underestimate the exposure prevalence, likely due to a 'healthy participant' bias,[36] reinforcing the need to use representative age- and sex-specific exposure prevalence data in PAF calculations.

We provide the first Australian estimates on the potential future burden of cancer avoidable through modification of current harmful exposures, using the latest available exposure prevalence estimates. Exposure to lifestyle-related risk factors is highly prevalent and largely increasing in Australia[23] and internationally,[5 37] and thus lifestyle modifications can have a large impact on the cancer burden. The exception to these trends is current smoking, as Australia is a world leader in smoking control, and prevalence rates are low and continuing to fall.[23 26 38] As we used the latest exposure prevalence data and evidence on cancers causally associated with specific exposures, recently updated for body fatness,[31] our PAF estimates are not directly comparable with previous Australian estimates.[8] These estimates were also based on published, mostly international, exposure-cancer associations, whereas we use harmonised Australian cohort data.

Accounting for competing risk of death is likely to have further increased the accuracy of our PAF estimates as ignoring competing risk of death can overestimate the fraction of cancers preventable by risk factor modification.[13] That is, if cancer and death share the same risk factors, reduction of these risk factors is likely to reduce the risk of cancer and the risk of death, and people living longer have increased opportunity to develop cancer. The bias is higher the more strongly the cancer risk factor is associated with death, the more risk factors are evaluated simultaneously and the longer the follow-up.[13] Our PAF method also produced CIs for the PAF estimates, allowing an evaluation of their precision and statistical comparison of subgroup estimates.

Our PAF method also allows a flexible choice of the reference level for a risk factor modification (eg, reducing the risk of current smokers to the level of former smokers) and analysis of the simultaneous effects of multiple risk factors. We showed that the combined contribution of two exposures on the cancer burden was overestimated if their effects on mortality were not accounted for and were assumed to be independent. We found that even a relatively small difference in PAF estimates can translate into a large difference in the number of preventable cancers predicted. Compared with the traditional method, our PAF estimates are thus more likely to reflect the real-world impact of modifying one or more risk factors[10] and to better inform future cancer control activities.

Limitations

Some risk factors were not collected by all studies, not available in the baseline data or the information available was too different for harmonisation, and as a result these studies could not be included in all analyses, reducing the statistical power. Additionally, some risk factors varied in how well they could be harmonised due to different question formulations and definitions (eg, 'daily' vs 'regular' smoking) or measurement methods (eg, self-reported vs measured BMI). Measurement error, both within and between studies, would generally lead to underestimation of the respective associations and PAF estimates. Additionally, as the exposure prevalence trends over time demonstrate, exposure to risk factors measured at baseline may have changed during follow-up, which would have further contributed to underestimation of the respective associations and PAFs. Some cohort studies performed repeated measurements during follow-up; these measures could be incorporated in future analyses as our PAF method allows the inclusion of time-dependent covariates.

Our illustrative PAF estimates for body fatness-related and tobacco and alcohol-related cancers were adjusted for age, sex and study and are thus subject to residual confounding by other risk factors affecting these associations. In the next stage of the project, we will compute cancer-specific PAF estimates and thoroughly evaluate and adjust for potential confounding factors.

The distribution of cancers in the cohort studies, especially when grouped, may not be the same as in the Australian population, and this may impact the generalisability of the findings. Reassuringly, the rank order of individual cancers in our cohort was similar to that for the Australian population.[2]

Our risk factor exposure prevalence estimates were obtained from study populations sampled to be representative of all Australians, but these surveys were limited in size and did not achieve 100% response rates, and therefore their representativeness is uncertain. For red and processed meat consumption, an important risk factor for several cancers, no prevalence information from such data sources was available; this further emphasises the importance of reaching a consensus on question formulations and definitions for core risk factors. Furthermore, even though we used the latest available exposure prevalence data, the estimates still lag behind the present situation. Therefore, our PAFs may be either slightly underestimated or overestimated, depending on the current exposure prevalence trends.

We note that probabilistic record linkage will have incurred a low rate of false positive and false negative matches, resulting in slight misclassification of the outcome.[39] Also, we were not able to capture loss to follow-up, for example, due to participants leaving Australia. Each of these limitations will likely have resulted in bias towards the null and PAF underestimation.[39]

Although we provide improved PAF estimates, further improvements are possible. One major assumption in the PAF estimation is an immediate reduction in risk after the hypothetical modification of the exposure of interest. This is unrealistic and therefore all PAF estimates overestimate the effect of the risk factor modification, or rather the time required for that effect to take place. Once reliable evidence on the lag time between an intervention and reduction in risk is available, it can be incorporated in the estimation of PAF using advanced modelling approaches.[34] The extent to which this is balanced by the various sources of underestimation mentioned above is not known and varies by risk factor. Furthermore, there is inherent uncertainty in the exposure prevalence estimates that could be incorporated in the PAF estimation for example through resampling-based methods such as bootstrap that require access to individual-level survey data. Finally, despite the large database available via our consortium, we have insufficient power to provide robust estimates for some of the rarer cancers, cancer subtypes, risk factor interactions and population subgroups. We aim to overcome this limitation by establishing an international cancer-PAF consortium.

Collaborators International cohort studies with risk factor, confounder and effect modifier, cancer and death data and access to representative up-to-date prevalence sources are welcome to participate. We encourage any interested parties to contact the corresponding author.

Author affiliations
[1]Centre for Big Data Research in Health, University of New South Wales, Sydney, Australia
[2]Cancer Research Division, Cancer Council New South Wales, Sydney, Australia
[3]School of Public Health, University of Sydney, Sydney, Australia
[4]Prince of Wales Clinical School, University of New South Wales, Sydney, Australia
[5]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia
[6]Diabetes and Population Health Laboratory, Baker IDI Heart and Diabetes Institute, Melbourne, Australia
[7]ANU College of Medicine, Biology and Environment, Australian National University, Canberra, Australia
[8]ANZAC Research Institute, University of Sydney and Concord Hospital, Sydney, Australia
[9]Research Centre for Gender, Health and Ageing, University of Newcastle, Newcastle, Australia
[10]School of Medicine, University of Adelaide, Adelaide, Australia
[11]Clinical Diabetes Laboratory, Baker IDI Heart and Diabetes Institute, Melbourne, Australia
[12]School of Nursing and Midwifery, University of South Australia, Adelaide, Australia
[13]School of Life and Environmental Sciences Charles Perkins Centre, University of Sydney, Sydney, Australia
[14]Centre for Vision Research, University of Sydney, Sydney, Australia

Acknowledgements The authors would like to thank the participating cohort studies and surveys and their participants for the data for this cohort consortium. Specific details of funding and data sources for the 45 and Up Study and the Australian Longitudinal Study on Women's Health are available at: www.saxinstitute.org.au/our-work/45-up-study/for-partners/ and www.alswh.org. The CHAMP study is funded by the National Health and Medical Research Council (ID301916) and the Ageing and Alzheimer's Institute. We acknowledge the assistance of the Data Linkage Unit at the Australian Institute of Health and Welfare for undertaking the data linkage to the ACD and the NDI.

Contributors MA, CMV and MAL were responsible for study concept, analysed and interpreted the data, drafted and revised the manuscript and approved the final version. KC, RM, PH, DJM, EB, GGG, RGC, JEB, AWT, JES, KP, VH, PM, B-AA contributed substantially to drafting the manuscript or revising it critically for important intellectual content and approved the final version.

Funding This study was funded by the National Health and Medical Research Council (ID1060991; ID1053642 to MAL; ID1082989 to KC; ID1042717 to EB) and a Cancer Institute New South Wales Fellowship (ID13/ECF/1-07 to MAL). Maria Arriaga was supported by Australian Postgraduate Award and a Translational Cancer Research Network (TCRN) PhD Scholarship Top-up Award. The funding bodies played no role in the conduct of the study, the writing of the report or the decision to submit the paper for publication.

Competing interests None declared.

Patient consent All the participating cohort studies have valid ethical approvals from human research ethics committees and have obtained informed consent for participation, collection and use of data for health research from each individual.

Ethics approval Our study has been approved by all necessary institutional and jurisdictional ethics committees: University of New South Wales human research ethics committee, Australian Institute of Health and Welfare ethics committee and each state and territory cancer registry human research ethics committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available. The data harmonisation guidelines can be obtained from the corresponding author. The SAS macros applied in the estimation of PAF are publicly available (see reference 15). We encourage international cohort studies interested in cohort-specific and pooled PAF analyses to contact the corresponding author.

REFERENCES

1. Australian Bureau of Statistics. *Causes of Death, Australia, 2015. 3303.0*. Canberra: ABS, 2016.
2. Australian Institute of Health and Welfare. *Cancer in Australia 2017. Cancer series No. 101. Cat. no. CAN 100*. Canberra: AIHW, 2017.
3. Levin ML. The occurrence of lung cancer in man. *Acta Unio Int Contra Cancrum* 1953;9:531–41.
4. Walter SD. The estimation and interpretation of attributable risk in health research. *Biometrics* 1976;32:829–49.
5. Forouzanfar MH, Alexander L, Anderson HR, *et al*. GBD 2013 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386:2287–323.
6. Whiteman DC, Wilson LF. The fractions of cancer attributable to modifiable factors: A global review. *Cancer Epidemiol* 2016;44:203–21.
7. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer* 2011;105 Suppl 2:S77–S81.
8. Whiteman DC, Webb PM, Green AC, *et al*. Cancers in Australia in 2010 attributable to modifiable factors: summary and conclusions. *Aust N Z J Public Health* 2015;39:477–84.
9. Australian Institute of Health and Welfare. *Risk factors contributing to chronic disease. Cat No. PHE 157*. Canberra: AIHW, 2012.
10. Geller K, Lippke S, Nigg CR. Future directions of multiple behavior change research. *J Behav Med* 2017;40:194–202.
11. Loef M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. *Prev Med* 2012;55:163–70.
12. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295–300.
13. Laaksonen MA, Härkänen T, Knekt P, *et al*. Estimation of population attributable fraction (PAF) for disease occurrence in a cohort study design. *Stat Med* 2010;29:860–74.
14. Steenland K, Armstrong B. An overview of methods for calculating the burden of disease due to specific risk factors. *Epidemiology* 2006;17:512–9.
15. Laaksonen MA, Virtala E, Knekt P, *et al*. *SAS* macros for calculation of population attributable fraction in a cohort study design. *J Stat Softw* 2011;43:1–25.
16. Giles GG, English DR. The melbourne collaborative cohort study. *IARC Sci Publ* 2002;156:69–70.
17. Mitchell P, Smith W, Attebo K, *et al*. Prevalence of open-angle glaucoma in Australia. The Blue Mountains Eye Study. *Ophthalmology* 1996;103:1661–9.
18. Lee C, Dobson AJ, Brown WJ, *et al*. Cohort Profile: the Australian longitudinal study on women's health. *Int J Epidemiol* 2005;34:987–91.
19. Dunstan DW, Zimmet PZ, Welborn TA, *et al*. Australian Diabetes, Obesity and Lifestyle Study (AusDiab). The Australian Diabetes, Obesity and Lifestyle Study (AusDiab)--methods and response rates. *Diabetes Res Clin Pract* 2002;57:119–29.
20. Grant JF, Taylor AW, Ruffin RE, *et al*. Cohort Profile: The North West Adelaide Health Study (NWAHS). *Int J Epidemiol* 2009;38:1479–86.
21. Cumming RG, Handelsman D, Seibel MJ, *et al*. Cohort Profile: the Concord Health and Ageing in Men Project (CHAMP). *Int J Epidemiol* 2009;38:374–8.
22. Banks E, Redman S, Jorm L, *et al*. Cohort profile: the 45 and up study. *Int J Epidemiol* 2008;37:941–7.
23. Australian Bureau of Statistics. *National Health Survey: First results, 2014-15. 4364.0.55.001*. Canberra: ABS, 2015.
24. Australian Bureau of Statistics. *Australian Health Survey: Summary of results, 2004-05. 4364.0*. Canberra: ABS, 2006.
25. Australian Bureau of Statistics. *Australian Health Survey: Summary of results, 2001. 4364.0*. Canberra: ABS, 2002.
26. Australian Institute of Health and Welfare. *National Drug Strategy Household Survey detailed report 2013. Drug statistics series no. 28. Cat. no. PHE 183*. Canberra: AIHW, 2014.

27. Velentzis LS, Banks E, Sitas F, *et al*. Use of menopausal hormone therapy and bioidentical hormone therapy in Australian women 50 to 69 years of age: results from a national, cross-sectional study. *PLoS One* 2016;11:e0146494.
28. Jaro MA. Probabilistic linkage of large public health data files. *Stat Med* 1995;14:491–8.
29. Friedman M. Piecewise Exponential Models for Survival Data with Covariates. *The Annals of Statistics* 1982;10:101–13.
30. International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risks to humans, Volume 100 (E). *A review of human carcinogens: Personal habits and indoor combustions*. Lyon, France: IARC, 2012:43-213, 373–501.
31. Lauby-Secretan B, Scoccianti C, Loomis D, *et al*. International Agency for Research on Cancer Handbook Working Group. Body Fatness and Cancer--Viewpoint of the IARC Working Group. *N Engl J Med* 2016;375:794–8.
32. Australian Institute of Health and Welfare Australian Cancer Database. 2012. http://www.aihw.gov.au/cancer-data/ (accessed Dec 2016).
33. Australian Institute of Health and Welfare. Cancer Series no. 66. Cat. No. CAN 62. *Cancer incidence projections: Australia, 2011 to 2020*. Canberra: AIHW, 2012.
34. Samuelsen SO, Eide GE. Attributable fractions with survival data. *Stat Med* 2008;27:1447–67.
35. Rolland B, Reid S, Stelling D, *et al*. Toward rigorous data harmonization in Cancer epidemiology research: one approach. *Am J Epidemiol* 2015;182:1033–8.
36. Sedgwick P. Bias in observational study designs: prospective cohort studies. *BMJ* 2014;349:g7731.
37. Vineis P, Wild CP. Global cancer patterns: causes and prevention. *Lancet* 2014;383:549–57.
38. National Preventative Health Taskforce. Tobacco control in Australia: making smoking history. Technical report No 2. *Australia: the healthiest country by 2020*. Canberra: Commonwealth of Australia, 2009.
39. Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. *Int J Epidemiol* 2002;31:1246–52.

BMJ Open

The burden of cancer attributable to modifiable risk factors: the Australian cancer-PAF cohort consortium

Maria E Arriaga, Claire M Vajdic, Karen Canfell, Robert MacInnis, Peter Hull, Dianna J Magliano, Emily Banks, Graham G Giles, Robert G Cumming, Julie E Byles, Anne W Taylor, Jonathan E Shaw, Kay Price, Vasant Hirani, Paul Mitchell, Barbara-Ann Adelstein and Maarit A Laaksonen

BMJ Open 2017 7:
doi: 10.1136/bmjopen-2017-016178