BMJ Open

Estimating the current and future cancer burden in Canada: methodological framework of the Canadian population attributable risk of cancer (ComPARe) study

Darren R Brenner,[1,2] Abbey E Poirier,[2] Stephen D Walter,[3] Will D King,[4] Eduardo L Franco,[5,6] Paul A Demers,[7] Paul J Villeneuve,[8] Yibing Ruan,[2] Farah Khandwala,[2] Xin Grevers,[2] Robert Nuttall,[9] Leah Smith,[9] Prithwish De,[10] Karena Volesky,[5,6] Dylan O'Sullivan,[4] Perry Hystad,[11] Christine M Friedenreich,[1,2] on behalf of the ComPARe Study Group

To cite: Brenner DR, Poirier AE, Walter SD, *et al*. Estimating the current and future cancer burden in Canada: methodological framework of the Canadian population attributable risk of cancer (ComPARe) study. *BMJ Open* 2018;**8**:e022378. doi:10.1136/bmjopen-2018-022378

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-022378).

Received 16 February 2018
Revised 1 June 2018
Accepted 29 June 2018

For numbered affiliations see end of article.

Correspondence to
Dr Darren R Brenner;
darren.brenner@ucalgary.ca

ABSTRACT

Introduction The Canadian Population Attributable Risk of Cancer project aims to quantify the number and proportion of cancer cases incident in Canada, now and projected to 2042, that could be prevented through changes in the prevalence of modifiable exposures associated with cancer. The broad risk factor categories of interest include tobacco, diet, energy imbalance, infectious diseases, hormonal therapies and environmental factors such as air pollution and residential radon.

Methods and analysis Using a national network, we will use population-attributable risks (PAR) and potential impact fractions (PIF) to model both attributable (current) and avoidable (future) cancers. The latency periods and the temporal relationships between exposures and cancer diagnoses will be accounted for in the analyses. For PAR estimates, historical exposure prevalence data and the most recent provincial and national cancer incidence data will be used. For PIF estimates, we will model alternative or 'counterfactual' distributions of cancer risk factor exposures to assess how cancer incidence could be reduced under different scenarios of population exposure, projecting incidence to 2042.

Dissemination The framework provided can be readily extended and applied to other populations or jurisdictions outside of Canada. An embedded knowledge translation and exchange component of this study with our Canadian Cancer Society partners will ensure that these findings are translated to cancer programmes and policies aimed at population-based cancer risk reduction strategies.

Strengths and limitations of this study

- We report a detailed and transparent approach for conducting large attributable risk estimation projects to assess the impact of multiple risk factors.
- We have considered projections of both the exposure prevalence and cancer incidence with multiple approaches, which is an improvement over unrealistic fixed projection models.
- Long-term projections of exposure prevalence and cancer incidence are statistically challenging and involve a great deal of uncertainty.
- Many of our exposure measures are based on self-reported data, which introduces the possibility of misreporting.

BACKGROUND

Estimates of the current and future burden of cancer in Canada attributable to known and probable causes of the disease are required for allocating prevention resources optimally. National[1 2] and global cancer incidence projections[3] suggest that the burden of cancer will continue to rise. In Canada and other developed nations, this is largely attributable to growing and ageing populations. In addition, despite established associations between modifiable risk factors and cancer risk, a sufficient reduction in the prevalence of these risk factors has not been achieved in Canada.[1–3] Identifying exposures and interventions with the greatest potential impacts of reducing cancer risk will aid in implementing prevention programmes and policies to combat this growing health challenge.

Several groups, including some members of our Canadian Burden of Cancer—Population Attributable Risk (ComPARe) Study Group, have produced estimates of the current burden of cancer attributable to lifestyle, environmental and infectious exposures in Canadian national[4–6] and provincial[7–15] populations. Additional studies have estimated the avoidable national[16–20] and

global[21] cancer burdens in future attributable to single exposures. However, population attributable risk (PAR) estimates are dependent on risk factor prevalence, which vary over time and are population specific. Therefore, it is important to frequently update PAR estimates. In addition, several methodological extensions to these approaches, including modelling the combined impact of multiple risk factors and defining the timing of intervention impacts on subsequent cancer incidence, are lacking. A comprehensive estimation of the current and future cancer burden and the impact of potential reductions in exposure prevalence on cancer incidence on the population are needed.

For the ComPARe Study, we developed a methodological framework to estimate the burden of cancer in Canada using cancer incidence data (2015) and projected incidence trends (2015–2042). The ComPARe study team brings together the substantive and quantitative expertise of cancer researchers from across the country. This collaborative, pan-Canadian study also involves a partnership with the Canadian Cancer Society, a main knowledge end-user for this work, who worked in partnership with the researchers throughout this project. To ensure methods were rigorously applied and standardised across research labs, we developed a methodological framework for the estimation of current attributable and future avoidable cancers associated with modifiable risk factors. This framework extends the work of other groups[22–29] and is applicable to a range of diseases and populations. In this paper, we describe the approach and methods used in the ComPARe Study. An overview of earlier methods used to estimate PARs and preventable impact fractions (PIFs) are presented. We then describe how we used these methods in the ComPARe Study, and the innovations that we developed to extend them. See figure 1 for an outline of our approach.

METHODS

Figure 1 shows the methodological framework for the ComPARe Study. The concept of PAR and population attributable fraction was initially developed by Levin in 1953 to estimate the burden of disease in the general population attributable to a given factor.[30] Attributable risks are predicated on the assumption that there are causal relationships between exposures and disease outcomes, and on the concept of the counterfactual, a scenario counter to actual experience, where exposures to the causal agents no longer exist or can be mitigated.[31]

Since the initial concept of the PAR method was introduced, several statistical and theoretical extensions to the framework have included methods to measure the uncertainty around PARs and the development of PIF. The PIF as an extension of PAR to consider situations of complete removal of the exposure cannot be assumed.[32] The impact of a reduction in the prevalence or population

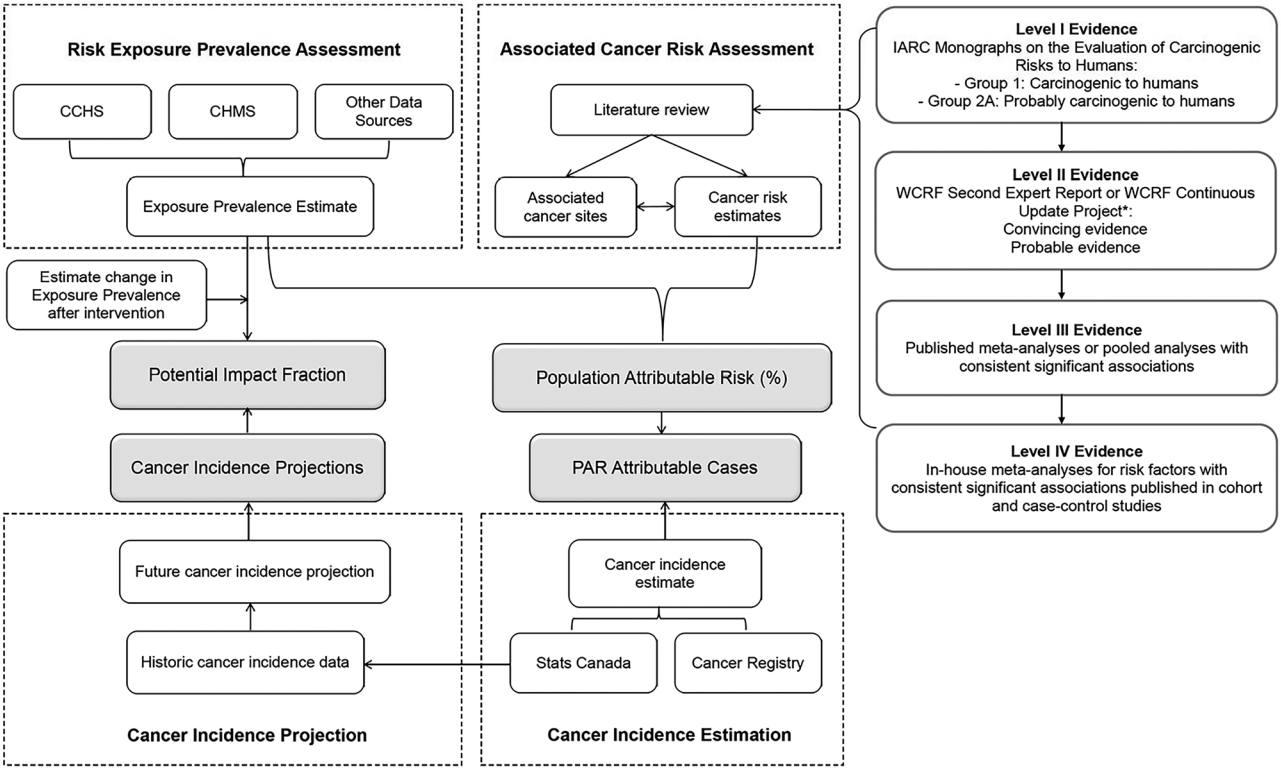

Figure 1 Scope of project framework for the estimation of current attributable and future avoidable disease burden. *Abbreviations*: CHMS, Canadian Health Measures Survey; CCHS, Canadian Community Health Survey; IARC, International Agency for Research on Cancer; WCRF, World Cancer Research Fund; CUP, continuous update project; PIF, potential impact fraction; PAR, population attributable risk.

Table 1 The population attributable risk estimation methods employed for the individual exposures of interest in the ComPARe project

Formula for PAR Estimation	Exposure
$Formula\ 1:\ PAR = \frac{Pe\ (RR-1)}{1+[Pe\ (RR-1)]}$	► Tobacco (secondhand smoke) ► UVR risk behaviours ► Disinfection by-products ► Low vitamin D ► Low dietary calcium intake ► *Helicobacter pylori* ► Hepatitis B ► Hepatitis C
$Formula\ 2:\ PAR = 1 - \sum_{i=0}^{k} \frac{p_{(c)i}}{RR_i}$	► Human papillomavirus ► Epstein–Barr virus ► Human T-cell lymphotropic virus type 1 ► Human herpes virus 8
$Formula\ 3:\ PAR = \frac{(p_{e1} \times ERR_1) + (P_{e2} \times ERR_2) + \ldots + (P_{ex} \times ERR_x)}{1+((p_{e1} \times ERR_1) + (P_{e2} \times ERR_2) + \ldots + (P_{ex} \times ERR_x))}$	► Tobacco (active exposure) ► Oral contraceptives ► Hormone replacement therapy ► Overweight/obesity ► Insufficient fruit and vegetable intake ► Red meat/processed meat intake ► High alcohol intake ► Insufficient dietary fibre intake ► Physical activity/inactivity
Individualised Methods	► Overall UV exposure ► Air pollution ► Radon ► Insufficient fruit and vegetable intake ► Red meat/processed meat intake ► Insufficient fibre intake ► Alcohol consumption

ERR, excess relative risk; i, exposure level; k, levels of exposure; PAR, population attributable risk; P_c, proportion of cases at the ith level of exposure; Pe, prevalence of exposure in the population; RR, relative risk; UV, ultraviolet; UVR, ultraviolet radiation.

distribution of an exposure and the subsequent impact of an exposure reduction is examined. The PAR and PIF form the statistical foundation of the ComPARe Study.

To apply the PAR and PIF to estimate the impacts of reducing exposures, three sources of data are essential (table 1): (1) the relative risk of incident disease, or risk distribution associated with exposure; (2) the proportion of the population or cancer cases exposed to the risk or protective factor (sex and age-specific exposure prevalence); and (3) sex-specific and age-specific disease incidence data. These three elements are needed to estimate the proportion of cancer cases that could be prevented, based on the PARs or PIFs. In the following sections, we present the methods used in the ComPARe Study for estimating the current attributable (PAR) and future avoidable (PIF) burdens of cancer.

Identifying risk factors for inclusion

A crucial component of attributable cancer estimation is determining which exposures should be included as causal for incident cancers. Given the considerable amount of epidemiological and basic science literature evaluating aetiological associations for cancer, we needed criteria to determine the level of evidence required for inclusion in our analyses. We developed a hierarchy of evidence for the ComPARe Study (figure 2) where quality determined using STrengthening the Reporting of OBservational studies in Epidemiology[33] guidelines for cohort and case-control studies and Meta-analysis Of Observational Studies in Epidemiology[34] guidelines for meta-analysis. The World Cancer Research Fund's (WCRF) Continuous Update Project[35] and the International Agency for Research on Cancer's (IARC) Monographs on the Evaluation of Carcinogenic Risks to Humans[36] have devoted substantial resources, including expert panels, to classifying potentially carcinogenic risks to humans. We used the recommendations from these international and national panels as our first level of inclusion. IARC group 1 (carcinogenic to humans) and group 2A (probably carcinogenic to humans) carcinogens were included. As a second level of evidence, we included exposure–cancer site pairs where high-quality meta-analyses of epidemiological studies published since the WCRF and IARC reports demonstrated consistent associations, as well as IARC Group 2B exposures for sensitivity analyses. The exposure and cancer site associations included in the ComPARe Study are presented in online supplementary table 1.

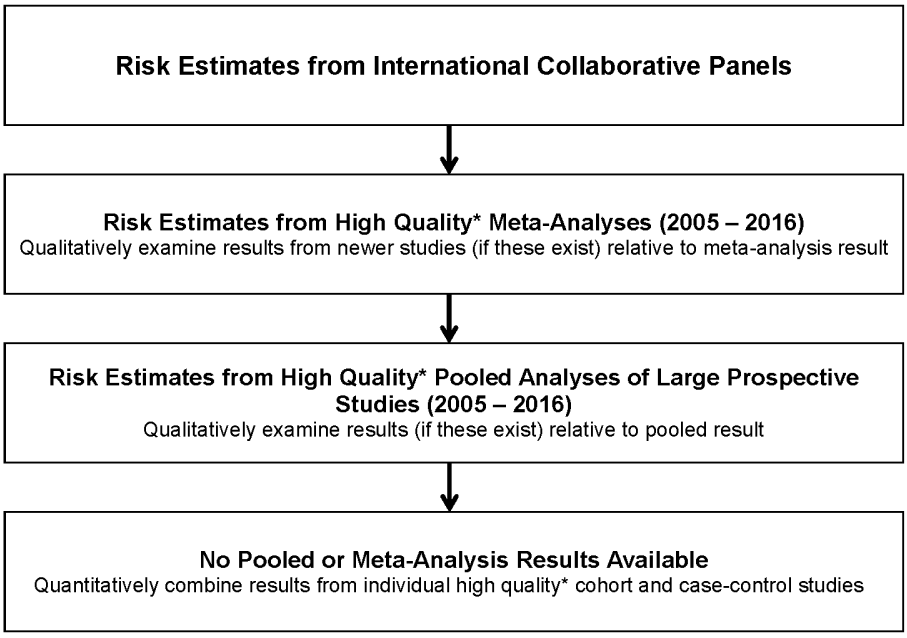

Figure 2 The process flow used for selecting risk estimates used in the ComPARe project. *Quality determined using STrengthening the Reporting of OBservational studies in Epidemiology[33] guidelines for cohort and case–control studies and Meta-analysis Of Observational Studies in Epidemiology[34] guidelines for meta-analysis.

Estimation of attributable cancers

Exposure prevalence data: including latency

The biologically relevant time period from the initiation of an exposure to development of disease is highly variable, depending on the exposure and cancer site, and it is likely to be measured in years or even decades for solid tumours. Therefore, we allowed for a period of latency from exposure to cancer incidence and diagnosis in our assessments. However, exposure prevalence data were not always available for the long relevant time periods implied by latency. As a proxy measure for each exposure, we extracted the median or mean follow-up time from exposure measurement to cancer incidence from large cohort studies. Our assessment of quality of the cohort studies was evaluated based on their sample size, methods of exposure assessment and length of follow-up, where large cohorts with detailed exposure and longer follow-up were considered the highest quality. This information concerning the latency period was then compared with the time periods for which high-quality data on exposure prevalence were available. We selected prevalence estimates that corresponded to the midpoint of the range of potential latency periods, as identified from the cohort studies. When these data were not available, we assumed a 10-year latency period between exposure measurement and cancer incidence, or used the closest available prevalence estimates. We attempted to strike a pragmatic balance between selecting a biologically plausible and relevant period of time and feasibly collecting prevalence data. For example, for the infectious agents, the latency period was determined by the *availability* of prevalence data. For *Helicobacter pylori*, there was one seroprevalence survey in 1999–2000, and for hepatitis B virus and hepatitis C virus, the prevalence data were collected from the Canadian Health Measures and the Canadian Notifiable Disease Surveillance System from 2007 to 2012. A diagram of our approach to modelling relevant exposures is shown in figure 3.

To estimate the attributable burden of cancer due to past exposures in Canada, we developed a hierarchy to select prevalence data from Canadian national and region-specific data sources, where available. For lifestyle exposures we considered data from large Canadian cohort studies when data from national population-based surveys were not available. For several environmental exposures, environmental monitoring data from sites in various parts of Canada were used. We collected exposure prevalence data overall and, where the data allowed, by sex, age and province.

Cancer incidence data

We obtained cancer incidence data for those 18 years of age and older from the Canadian Cancer Registry

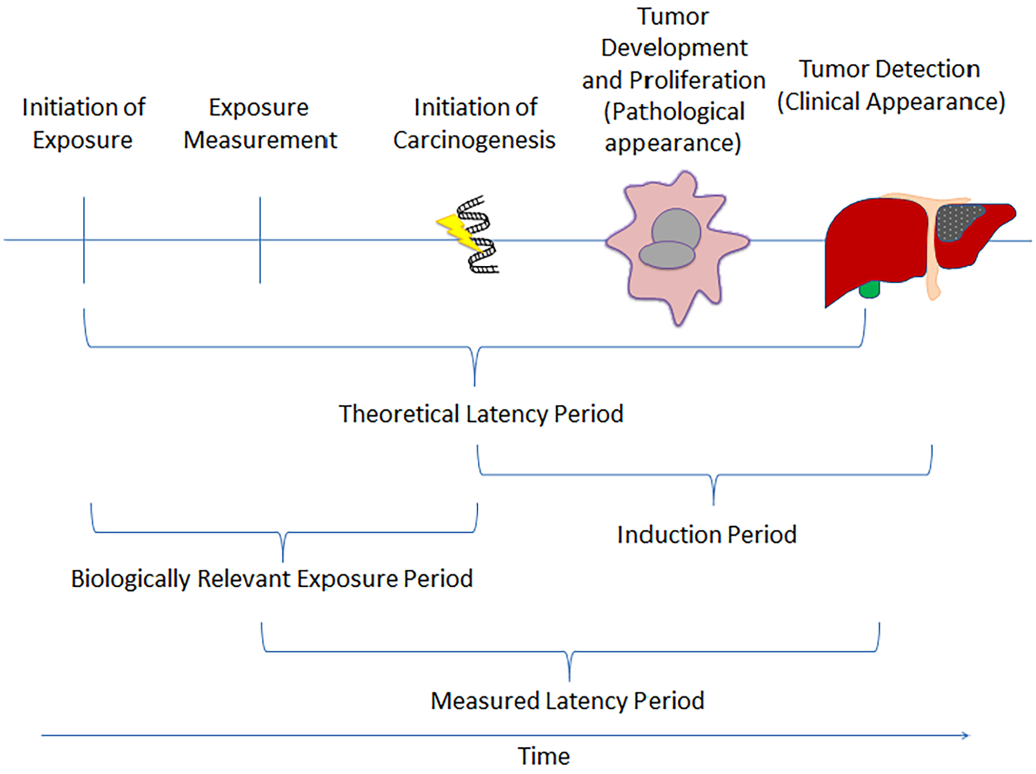

Figure 3 Representation of relevant exposure windows and latency onset considered for the ComPARe project.

(CCR), a national registry of cancer cases covering the entire population of Canada, including by province and territory. Statistics Canada produces annual data quality reports for the CCR and each Canadian province and territory has a legislated responsibility for cancer collection and control, which improves the completeness and population coverage of the data.[37] Data by province, sex and 5-year age group for 2012, being the most recent year of national data available at the time of the study (except for Quebec data, which were extrapolated from 2010), were obtained. Cancer cases were coded in the CCR using the International Classification of Diseases for Oncology, 3rd Edition. Cancer mortality was not considered in this study as we were interested in cancer prevention through changes in behaviours and exposures. Furthermore, the inclusion of survival requires an additional set of modelling assumptions related to survival across exposures groups, where the evidence base is far less developed.

Estimation of population attributable cancers—including uncertainty

The PAR estimation methods employed for the individual exposures in the ComPARe Study are presented in table 1. Since 95% CIs cannot be easily calculated for PARs,[38] Monte Carlo simulation methods were used to estimate 95% CIs around PAR estimates, where the relative risk (RR) values were drawn from a log normal distribution derived from the RR and its associated variance estimated from 95% CIs while prevalence values were drawn from a binomial distribution with parameter n as the number of survey participants and parameter p as the prevalence of exposure estimated from the survey. We simulated 10 000 samples and used the 2.5th and 97.5th percentiles of the resulting PAR distribution as the lower and upper limits of its 95% CI.[39 40]

Estimation of avoidable cancers

Exposure prevalence data

To estimate the future avoidable cancer burden to 2042, it is necessary to project exposure prevalence (eg, to 2032 if a 10-year latency period is used). We used the exposure prevalence data hierarchy outlined above to identify the optimal exposure prevalence data. For these data, we focused on sources with longitudinal surveys. For exposures where historical data allowed past trends to be observed, one of the several approaches to model future prevalence were used. These included linear, logistic growth, multinomial logistic regression and exponential curves to predict the future proportion of the population exposed. Prevalence estimates were projected by sex and various levels of exposure prevalence. Models were selected based on expert opinion of the visual evaluation of the fit to past data trends and by avoiding extreme projection scenarios that might have arisen because of some overly influential data points. The different approaches to model future prevalence reflect different potential scenarios. Logistic growth considers that the prevalence of exposure would reach a future steady state, while multinomial logistic regression predicts that the past exposure observed trend would continue relatively unchanged into the future. Exponential and logarithmic curves are a compromise between the logistic and multinomial approaches, and involve an assumption that the past trend would continue, but at a slower pace. We

projected exposure data for the combined population and for males and females separately, for both national and provincial estimates, where the data allowed.

Cancer incidence projections

Cancer incidence frequencies and rates were projected by extrapolating past trends using various statistical models. In the past, trends over age at diagnosis, year of diagnosis (period) and/or year of birth (cohort) as well as hybrids of these models have been used. More recently, the age-period-cohort[41] and the age-drift-period-cohort (Nordpred)[42] models have been widely used. For the ComPARe Study, the R package 'Canproj'[43] was used to project cancer incidence from 2012 to 2042. The package projects forward to a maximum of 30 years, which suited our needs, based on the uncertainty surrounding cancer sites for which secondary or primary prevention interventions were being scaled up (eg, colorectal, breast, lung and cervical cancers) or reduced (eg, prostate cancer).

Canproj combines cancer projection methods that have been used in the last 30 years to select the best fitted model for the data, using a decision algorithm to identify the most appropriate projection (online supplementary figure 1). The models available in Canproj include age-only, age-period (including common trend and age-specific trend), age-cohort and Nordpred[42] (age-drift-period-cohort; negative-binomial distribution may replace the Poisson distribution when over-dispersion appears). All models provide projected age-specific incidence rates and counts. Through the decision algorithm the Canproj methods produce more realistic projection estimates than other approaches, such as the Poisson regression method,[44] the polynomial regression and natural spline methods,[45] the joinpoint method[46] and the Bayesian Markov chain Monte Carlo methods[47] by taking advantage of specific aspects of all these methods to fit the best model. We evaluated all the findings, independently of goodness-of-fit to inspect the face validity of the projections.

Defining counterfactual scenarios

Within our avoidable cancers (PIF) framework we examined a range of exposure prevalence reduction scenarios or counterfactuals. Our primary counterfactuals were based on population-based interventions that have been shown to be beneficial in experimental studies, and which could be scaled up to the population level. We conducted a systematic literature search of interventions for each exposure and identified their effects from reviews, meta-analyses or large intervention (individual and/or community level) trials. For all exposures, we also included models with fixed prevalence reductions of 10%, 25%, 50% and 100% for every year between 2018 and 2042.

Potential impact fraction estimation: defining latency of interventions

Using projected exposure prevalence, cancer incidence and a range of counterfactual scenarios, we then estimated the proportions and numbers of avoidable cancers in Canada from 2018 to 2042. To present these results, we plotted the number of projected cancers under the baseline projection scenario (if no change in exposure prevalence were to occur), followed by the incidence estimated under a range of counterfactual scenarios.

To evaluate the assumed fixed latency period, we conducted sensitivity analyses using some other assumptions for the statistical distribution of latency periods, for example, including the uniform, modified Weibull and binomial distributions. These alternative distributions were each chosen to have a mean of 10 years and range from 0 to 15 years. Incorporating a distribution of latency periods into PIF estimation allowed us to better predict the transitional effect of counterfactual interventions.

Consideration of multiple risk factors and joint effects

As with other burden estimation efforts, our primary analyses were focused on the attributable and avoidable proportions and numbers of cancers related to individual exposures separately. This approach is an oversimplification because several exposures might be known to have joint impact or interactions on cancer risk. Several well-characterised examples include alcohol and tobacco for various cancer sites,[48] and overweight or obesity and physical inactivity for colorectal cancer.[49] Where possible, we have also estimated the impact of multiple risk factors for a series of scenarios where the scientific literature has suggested the existence of combined or synergistic effects. When exposures are strongly associated and/or their interaction on cancer risk departs from multiplicative risk, Levin's formula to estimate PAR of individual risk factors must be used with caution. In order to combine PAR across exposures we used the Miettinen-Steenland approach for any combined or 'summary' estimates.

Sensitivity analyses

Our sensitivity analyses sought to characterise potential bias in the available prevalence and risk data. Since we relied on data from self-report questionnaires for some exposures, such as alcohol, physical activity and body weight, we expected a certain degree of misreporting. In our sensitivity analyses, we corrected the reported prevalence by using studies that had validated the survey data, based on small samples of objective measurements, and then using sex-specific correction factors. Some exposures had considerable (>10%) non-response rates (ie, responded 'don't know' or 'refuse to answer'), and for these cases in our main analysis, we assumed that non-responders had been unexposed to the risk factors in question. In the sensitivity analyses, we imputed exposure values using both missing-at-random and missing-not-at-random assumptions. For the missing-at-random scenario, we assumed that non-response was unrelated to the exposure status, and hence that the exposure distribution among non-responders was identical to that of responders. For the missing-not-at-random scenario, we assumed that the non-responders were all exposed, and that their exposure distribution was identical to the exposed survey responders.

Patient and public involvement

No patients or public were involved in this study protocol.

DISCUSSION

In the ComPARe Study, we developed approaches for each step of data collection, analysis, uncertainty estimation and sensitivity analyses in order to arrive at plausible PAR estimates for cancer incidence. Furthermore, this approach provides a methodologically rigorous framework for long-term projections of cancer burden and the relative impacts of different population-based interventions for cancer prevention. As new cancer risk factor prevention strategies are developed, their subsequent impact on the future cancer burden can easily be integrated into this project for a comparative analysis of intervention strategies.

The estimates from this project will be relevant to a broad audience, ranging from those working in cancer prevention and more broadly in health promotion, to cancer advocacy groups, public health and healthcare planners, health policymakers, clinicians and the public to inform priority setting in prevention programming and resources; allocation of funding to areas of unmet need; and so on. We have developed this project in collaboration with the Canadian Cancer Society (CCS), our knowledge translation partner. As a primary end-user of the data generated from this project, CCS's input into the design and desired output of the project has been invaluable. We encourage other groups to plan knowledge translation via similar partnership arrangements from the initiation phase of the project.

Methodological extensions

During this project we encountered several methodological components that were comparatively under-developed. For example, while several groups have conducted large attributable risk estimation projects, few, if any, have systematically assessed the impact of multiple risk factors. Our examination of approaches for multiple risk factors adds to the literature and provides validation of the estimates produced in this project. In addition, we have considered projections of both the exposure prevalence and cancer incidence data with multiple approaches. Previous projects have assumed fixed cancer incidence or exposure prevalence for future projections, and both are unrealistic. Furthermore, in the application of our counterfactual scenarios, we tested and applied several lag time models to fit the most likely windows of exposure and their associated subsequent changes in cancer incidence. In addition, we have worked in collaboration with key knowledge end-users to develop counterfactual scenarios that best match realistic expectations for cancer prevention programmes.

Limitations

Our framework, while building on previous approaches, has a number of limitations. Long-term projections of exposure prevalence and cancer incidence are statistically challenging and involve a great deal of uncertainty. Although we have strived to identify the highest quality exposure prevalence and cancer incidence datasets, and used methodologically sound approaches for modelling, our results still need to be interpreted with caution. The resulting projections are a direct product of the validity of the input data on exposure prevalence and associated relative risks. Using data of poor quality or having questionable validity may result in erroneous projections. For this reason, we included population-based, nationally representative surveys to estimate exposure prevalence when they were available. Many of our exposure measures, particularly for the lifestyle risk and protective factors, were based on self-reported data. Where possible, we modelled the potential impact of reporting biases on our estimates and included analyses focused on directly measured exposures.

For several infectious agents including Epstein-Barr virus, *H. pylori* and human papillomavirus, large-population-based estimates of prevalence were not available for Canada. For these instances, we included case series, case–control and cohort studies, as well as population-based surveys extracted from populations from the USA and if not available, then Western Europe. The use of a more sensitive assay for the detection of *H. pylori* has substantially increased the proportion of non-cardiac gastric cancers attributable to this infectious agent.[50] To account for the new gold standard, the included studies will be corrected for measurement error.

In terms of cancer incidence projections, we relied on the Canproj programme,[43] which uses age-period-cohort models and the extension of the Nordpred model that has been widely used by other research groups for long-term projections of cancer incidence. However, errors in estimates are inevitable when projecting 30 years into the future as the models do not account for future changes in risk factors (ie, population changes in smoking patterns, diet, etc.). In addition, to deal with some of the uncertainty inherent in projections, expert opinion was used when the projection model selected by Canproj was implausible, which introduces some degree of bias to the decisions.

The CCR is a high-quality database with good case ascertainment of malignant tumours. Very few incident cancer cases are missed in the CCR and therefore any bias would be minimal and would not affect our results.[37] However, data for the province of Quebec were extrapolated from 2010, as data for 2012 were not available, which is a limitation for the national counts. Ethnicity was not taken into account in these estimates for various reasons. Unlike other national cancer registries, the CCR does not provide incidence data by ethnicity. Canada is not a populous country and stratifying cancer incidence by sex, age and ethnicity would lead to few observations. Furthermore, ethnicity-specific risk estimates and prevalence data are not available at this time. However, for ultraviolet radiation (UVR) exposure, ethnicity was taken

into account, as there is a strong interaction between UVR and ethnicity.

CONCLUSIONS

We have described a methodological framework for attributable risk estimation and cancer projection that extends our previous research in PAR and PIFs. The application of this framework will provide estimates of both current attributable and future avoidable disease risk in Canada. These findings will be of use to those working in cancer prevention, public health and healthcare planners, health policymakers, healthcare providers and the general public for a wide range of applications in cancer control and prevention.

Author affiliations
[1]Departments of Oncology and Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[2]Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada
[3]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
[4]Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada
[5]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada
[6]Department of Oncology, McGill University, Montreal, Quebec, Canada
[7]Occupational Cancer Research Centre, Cancer Care Ontario, Toronto, Ontario, Canada
[8]Department of Health Sciences, Carleton University, Ottawa, Ontario, Canada
[9]Canadian Cancer Society, Toronto, Ontario, Canada
[10]Cancer Care Ontario, Toronto, Ontario, Canada
[11]College of Public Health and Human Sciences, Oregon State University, Corvallis, Oregon, USA

Collaborators Elizabeth Holmes (Canadian Cancer Society, Toronto, Ontario, Canada), Zeinab El-Masri (Cancer Care Ontario, Toronto, Ontario, Canada), Mariam El-Zein, Sheila Bouten (Department of Oncology, McGill University, Montréal, Québec, Canada), Tasha Narain, Priyanka Gogna (Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada).

Contributors DRB, CMF, SDW, WDK, ELF, PAD, PJV, RN, and PD. were responsible for the study conception. DRB, AEP, SDW, WDK, ELF, PAD, PJV, YR, FK, XG, RN, LS, PD, KV, DO'S, PH and CMF contributed substantially to the study design. DRB, AEP and YR drafted the manuscript. DRB, AEP, SDW, WDK, ELF, PAD, PJV, YR, FK, XG, RN, LS, PD, KV, DO'S, PH and CMF revised the draft paper, and gave final approval of this version to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This research is supported by a Canadian Cancer Society Research Institute Partner Prevention Research Grant (#703106).

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Please contact the corresponding author for inquiries related to the data resulting from this work.

REFERENCES

1. Canadian Cancer Society's Advisory Committee on Cancer Statistics. *Canadian Cancer Statistics 2015*. Toronto, Ontario: Canadian Cancer Society, 2015.
2. Xie L, Semenciw R, Mery L. Cancer incidence in Canada: trends and projections (1983-2032). *Health Promot Chronic Dis Prev Can* 2015;35(Suppl 1):2–186.
3. Bray F, Møller B. Predicting the future burden of cancer. *Nat Rev Cancer* 2006;6:63–74.
4. Chen J, Moir D, Whyte J. Canadian population risk of radon induced lung cancer: a re-assessment based on the recent cross-Canada radon survey. *Radiat Prot Dosimetry* 2012;152:9–13.
5. Al-Arydah M. Population attributable risk associated with lung cancer induced by residential radon in Canada: Sensitivity to relative risk model and radon probability density function choices: In memory of Professor Jan M. Zielinski. *Sci Total Environ* 2017;596-597:331–41.
6. Krueger H, Turner D, Krueger J, *et al*. The economic benefits of risk factor reduction in Canada: tobacco smoking, excess weight and physical inactivity. *Can J Public Health* 2014;105:69–78.
7. Krueger H, Koot JM, Rasali DP, *et al*. Regional variations in the economic burden attributable to excess weight, physical inactivity and tobacco smoking across British Columbia. *Health Promot Chronic Dis Prev Can* 2016;36:76–86.
8. Peterson E, Aker A, Kim J, *et al*. Lung cancer risk from radon in Ontario, Canada: how many lung cancers can we prevent? *Cancer Causes Control* 2013;24:2013–20.
9. Brenner DR, Poirier AE, Grundy A, *et al*. Cancer incidence attributable to inadequate physical activity in Alberta in 2012. *CMAJ Open* 2017;5:E338–44.
10. Brenner DR, Poirier AE, Grundy A, *et al*. Cancer incidence attributable to excess body weight in Alberta in 2012. *CMAJ Open* 2017;5:E330–6.
11. Grevers X, Grundy A, Poirier AE, *et al*. Cancer incidence attributable to the use of oral contraceptives and hormone therapy in Alberta in 2012. *CMAJ Open* 2016;4:E754–9.
12. Grundy A, Poirier AE, Khandwala F, *et al*. Cancer incidence attributable to red and processed meat consumption in Alberta in 2012. *CMAJ Open* 2016;4:E768–75.
13. Grundy A, Poirier AE, Khandwala F, *et al*. Cancer incidence attributable to insufficient fruit and vegetable consumption in Alberta in 2012. *CMAJ Open* 2016;4:E760–7.
14. Grundy A, Poirier AE, Khandwala F, *et al*. Cancer incidence attributable to insufficient fibre consumption in Alberta in 2012. *CMAJ Open* 2017;5:E7–E13.
15. Poirier AE, Grundy A, Khandwala F, *et al*. Cancer incidence attributable to tobacco in Alberta, Canada, in 2012. *CMAJ Open* 2016;4:E578–87.
16. Kuhn R, Rothman DS, Turner S, *et al*. Beyond Attributable Burden: Estimating the Avoidable Burden of Disease Associated with Household Air Pollution. *PLoS One* 2016;11:e0149669.
17. Baade PD, Meng X, Sinclair C, *et al*. Estimating the future burden of cancers preventable by better diet and physical activity in Australia. *Med J Aust* 2012;196:337–40.
18. Pandeya N, Wilson LF, Webb PM, *et al*. Cancers in Australia in 2010 attributable to the consumption of alcohol. *Aust N Z J Public Health* 2015;39:408–13.
19. Kendall BJ, Wilson LF, Olsen CM, *et al*. Cancers in Australia in 2010 attributable to overweight and obesity. *Aust N Z J Public Health* 2015;39:452–7.
20. Nagle CM, Wilson LF, Hughes MC, *et al*. Cancers in Australia in 2010 attributable to the consumption of red and processed meat. *Aust N Z J Public Health* 2015;39:429–33.
21. Arnold M, Pandeya N, Byrnes G, *et al*. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. *Lancet Oncol* 2015;16:36–46.
22. Whiteman DC, Webb PM, Green AC, *et al*. Cancers in Australia in 2010 attributable to modifiable factors: introduction and overview. *Aust N Z J Public Health* 2015;39:403–7.
23. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3:e442.
24. World Health Organization. *Global health risks: mortality and burden of disease attributable to selected major risks*. Geneva: World Health Organization, 2009.
25. Parkin DM. 1. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer* 2011;105(S2):S2–S5.
26. Barendregt JJ, Veerman JL. Categorical versus continuous risk factors and the calculation of potential impact fractions. *J Epidemiol Community Health* 2010;64:209–12.

27. Arnold M, Touillaud M, Dossus L, *et al*. Cancers in France in 2015 attributable to high body mass index. *Cancer Epidemiol* 2018;52:15–19.
28. Menvielle G, Kulhánová I, Bryère J, *et al*. Tobacco-attributable burden of cancer according to socioeconomic position in France. *Int J Cancer* 2018;143:478–85.
29. Shield KD, Marant Micallef C, Hill C, *et al*. New cancer cases in France in 2015 attributable to different levels of alcohol consumption. *Addiction* 2018;113:247–56.
30. Levin ML. The occurrence of lung cancer in man. *Acta Unio Int Contra Cancrum* 1953;9:531–41.
31. Morgan SL, Winship C. *Counterfactuals and causal inference*. Cambridge: Cambridge University Press, 2014.
32. Morgenstern H, Bursic ES. A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. *J Community Health* 1982;7:292–309.
33. von Elm E, Altman DG, Egger M, *et al*. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806–8.
34. Stroup DF, Berlin JA, Morton SC, *et al*. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Jama* 2000;283:2008–12.
35. World Cancer Research Fund / American Institute for Cancer Research. *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective*. Washington, D.C: American Institute for Cancer Research, 2007.
36. International Agency for Reseach on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Lyon, France. 2017 http://monographs.iarc.fr/ (accessed 29 June 2017).
37. Statistics Canada. *Canadian Cancer Registry (CCR)*. Ottawa, ON: Statistics Canada, 2014.
38. Greenland S. Interval estimation by simulation as an alternative to and extension of confidence intervals. *Int J Epidemiol* 2004;33:1389–97.
39. Lee IM, Shiroma EJ, Lobelo F, *et al*. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* 2012;380:219–29.
40. Renehan AG, Soerjomataram I, Tyson M, *et al*. Incident cancer burden attributable to excess body mass index in 30 European countries. *Int J Cancer* 2010;126:692–702.
41. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat Med* 1987;6:469–81.
42. Møller B, Fekjaer H, Hakulinen T, *et al*. Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Cancer Prev* 2002;11 Suppl 1–S1–S96.
43. Qiu Z, Hatcher J, Team C-PW. *Canproj—The R package of cancer projection methods based on generalized linear models for age, period, and/or cohort*. Alberta Health Services: Technique Report for Cancer Projections Network (C-Proj) Alberta, 2011.
44. Dyba T, Hakulinen T, Päivärinta L. A simple non-linear model in incidence prediction. *Stat Med* 1997;16:2297–309.
45. Carstensen B. Age-period-cohort models for the Lexis diagram. *Stat Med* 2007;26:3018–45.
46. Kim HJ, Fay MP, Feuer EJ, *et al*. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med* 2000;19:335–51.
47. Bray I. Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *Journal of the Royal Statistical Society: Series C* 2002;51:151–64.
48. Hart CL, Davey Smith G, Gruer L, *et al*. The combined effect of smoking tobacco and drinking alcohol on cause-specific mortality: a 30 year cohort study. *BMC Public Health* 2010;10:789.
49. Samad AK, Taylor RS, Marshall T, *et al*. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Dis* 2005;7:204–13.
50. Plummer M, Franceschi S, Vignat J, *et al*. Global burden of gastric cancer attributable to Helicobacter pylori. *Int J Cancer* 2015;136:487–90.