BMJ Open

Estimating the Current and Future Cancer Burden in Canada: Methodologic Framework of the Canadian Population Attributable Risk of Cancer (ComPARe) Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-022378
Article Type:	Protocol
Date Submitted by the Author:	16-Feb-2018
Keywords:	Epidemiology < ONCOLOGY, Potential impact fraction, Population attributable risk, Cancer, STATISTICS & RESEARCH METHODS, Risk factors

Complete List of Authors:
Brenner, Darren; University of Calgary Cumming School of Medicine, Oncology; Alberta Health Services, Cancer Epidemiology and Prevention Research
Poirier, Abbey; Alberta Health Services, Cancer Epidemiology and Prevention Research
Walter, Stephen ; McMaster University , Department of Health Research Methods, Evidence, and Impact
King, Will; Queen's University, Department of Public Health Sciences
Franco, Eduardo; McGill University, Department of Oncology; McGill University Department of Epidemiology Biostatistics and Occupational Health
Demers, Paul; Cancer Care Ontario, Occupational Cancer Research Centre
Villeneuve, Paul; Carleton University Department of Health Sciences
Ruan, Yibing; Alberta Health Services, Cancer Epidemiology and Prevention Research
Khandwala, Farah; Alberta Health Services, Cancer Epidemiology and Prevention Research
Grevers, Xin; Alberta Health Services, Cancer Epidemiology and Prevention Research
Nuttall, Robert; Health Quality Ontario
Smith, Leah; Canadian Cancer Society
De, Prithwish; Cancer Care Ontario
Volesky, Karena; McGill University, Department of Oncology; McGill University Department of Epidemiology Biostatistics and Occupational Health
O'Sullivan, Dylan; Queen's University, Department of Public Health Sciences
Hystad, Perry ; Oregon State University
Friedenreich, CM; Alberta Health Services, Cancer Epidemiology and Prevention Research; University of Calgary Cumming School of Medicine, Oncology

Estimating the Current and Future Cancer Burden in Canada: Methodologic 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 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[12]

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, Montréal, Québec, Canada
6. Department of Oncology, McGill University, Montréal, Québec, Canada
7. Occupational Cancer Research Centre, 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, United States
12. Additional members of the ComPARe study group:
    Elizabeth Holmes[9]
    Zeinab El-Masri[10]
    Mariam El-Zein[6]
    Sheila Bouten[6]
    Tasha Narain[4]
    Priyanka Gogna[4]

*To Whom Correspondence should be addressed:
Darren R. Brenner
Department of Cancer Epidemiology and Prevention Research
CancerControl Alberta, Alberta Health Services
Holy Cross Centre – Room 513C
Box ACB, 2210-2nd St. SW, Calgary, AB
T2S 3C3
Phone: 1-403-698-8178
Darren.Brenner@ucalgary.ca

ABSTRACT

Introduction: The Canadian Population Attributable Risk of Cancer (ComPARe) project will quantify the number and proportion of incident cancer cases 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: Working as 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 for 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 programs 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 cancer are required to allocate 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 aging populations. In addition, despite established associations between modifiable risk factors and cancer risk, sufficient reductions in the prevalence of these risk factors have 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 programs 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 future avoidable national [16-20] and global [21] cancer burdens 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 methodologic extensions to these approaches, including modeling 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 of the impact of potential reductions in exposure prevalence on cancer incidence at 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 standardized 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 [22-26] and is applicable to a range of diseases and populations. Here 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 methodologic framework for the ComPARe study. The concept of PAR or population attributable fraction (PAF) was initially developed by Levin in 1953 to estimate the burden of disease in the general population attributable to a given factor.[27] 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.[28]

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 the potential impact fraction (PIF). The PIF is an extension of the PAR to consider situations complete removal of the exposure cannot be assumed.[29] The impact of a reduction in the prevalence or population distribution of an exposure and the subsequent impact of an exposure reduction is examined. The PAR and PIF are 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 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 epidemiologic and basic science literature evaluating etiologic 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). The World Cancer Research Fund's (WCRF) Continuous Update Project [30] and the International Agency for Research on Cancer's (IARC) Monographs on the Evaluation of Carcinogenic Risks to Humans[31] 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. As a second level of evidence, we included exposure/cancer site pairs where high quality meta-analyses of epidemiologic studies published since the WCRF and IARC reports demonstrated consistent associations. The exposure and cancer site associations included in the ComPARe Study are presented in Supplementary Table 1.

Estimation of Attributable Cancers

Exposure Prevalence Data – Including Latency

The biologically relevant time period from 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 tumors. Therefore, we allowed for a period of latency from exposure to cancer incidence/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. 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 (CCR); a national registry of cancer cases covering the entire population of Canada, including by province and territory. Statistisc 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[32]. Data by province, sex and five-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 (ICD-O-3).

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% confidence intervals (CIs) cannot be easily calculated for PARs[33], Monte Carlo simulation methods were used to estimate 95% CIs around PAR estimates, where the 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.[34 35]

Estimation of Avoidable Cancers

Exposure Prevalence Data

To estimate the future avoidable cancer burden to 2042, it is necessary to project exposure prevalence (e.g. 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 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 the 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 /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 [36] and the age-drift-period-cohort (Nordpred) [37] models have been widely used. For the ComPARe study, the R package 'Canproj' [38] 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 (e.g., colorectal, breast, lung, and cervical cancers) or reduced (e.g., 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 (Supplementary Figure 1). The models available in Canproj include: age-only, age-period (including common trend and age-specific trend), age-cohort and Nordpred [37] (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 [39], the polynomial regression and natural spline methods [40], the joinpoint method [41] and the Bayesian Markov Chain Monte Carlo methods [42] by taking advantage of specific aspects of all of these methods to fit the best model. We evaluated all 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 which 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, e.g., 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/interactions on cancer risk. Several well-characterized examples include alcohol and tobacco for various cancer sites [43], and overweight/obesity and physical inactivity for colorectal cancer.[44] 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.. We also compared different approaches for estimating the combined PARs for multiple risk factors. For example, under certain situations, summing the PARs for each exposure can give an approximate estimate of their combined PAR. Details regarding these methods are presented in a separate publication (Ruan, 2017).

Sensitivity analyses

Our sensitivity analyses sought to characterize 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 (i.e. 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.

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 policy makers, clinicians and the public to inform

priority setting in prevention programming and resources; allocation of funding to areas of unmet need; etc. We have developed this project in collaboration with our knowledge translation partner – the Canadian Cancer Society (CCS). 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.

Methodologic Extensions

During this project we encountered several methodologic 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, 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 programs.

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/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, *Helicobacter 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 United States and if not available, then Western Europe. For *Helicobacter pylori*, we corrected the measurement error present in relative risk estimates on the basis of a more sensitive assay to measure seroprevalence.

In terms of cancer incidence projections, we relied on the Canproj program [38], 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 in the future as the models do not account for future changes in risk factors (i.e. population changes in smoking patterns, diet, etc.). The Canadian Cancer Registry 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.[32] 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.

Conclusions

We have described a methodologic 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 policy makers, healthcare providers and the general public for a wide range of applications in cancer control and prevention.

Competing Interests

We have read and understood the BMJ Group policy on declaration of interests and declare no competing interests.

Contributor Statement

D.R.B, C.M.F., S.D.W., W.D.K, E.L.F., P.D., P.J.V., R.N., and P.De. were responsible for the study conception. D.R.B., A.E.P., S.D.W., W.D.K., E.L.F., P.D., P.J.V., Y.R., F.K., X.G., R.N, L.S., P.De., K.V., D.O., P.H, C.M.F. contributed substantially to the study design. D.R.B, A.E.P. and Y.R. drafted the manuscript. D.R.B., A.E.P., S.D.W., W.D.K., E.L.F., P.D., P.J.V., Y.R., F.K., X.G., R.N, L.S., P.De., K.V., D.O., P.H, C.M.F. revised the draft paper, 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.

Licence to BMJ Publishing Group Limited

I Darren Brenner, The Corresponding Author of this article contained within the original manuscript which includes any diagrams & photographs within and any related or stand-alone film submitted has the right to grant on behalf of all authors and does grant on behalf of all authors, a licence to the BMJ Publishing Group Ltd and its licencees, to permit this Contribution (if accepted) to be published in the BMJ Open and any other BMJ Group products and to exploit all subsidiary rights.

Transparency Declaration

Darren Brenner affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Funding

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

References

1. Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2015.Special topic: Predictions of the future burden of cancer in Canada. Toronto, Ontario: Canadian Cancer Society, 2015.

2. Xie L, Semenciw R, Mery L. Cancer incidence in Canada: trends and projections (1983-2032). *Health promotion and chronic disease prevention in Canada : research, policy and practice* 2015;35 Suppl 1:2-186. [published Online First: 2015/05/27]

3. Bray F, Moller B. Predicting the future burden of cancer. *Nat Rev Cancer* 2006;6(1):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. *Radiation protection dosimetry* 2012;152(1-3):9-13. doi: 10.1093/rpd/ncs147 [published Online First: 2012/08/10]

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. *The Science of the total environment* 2017;596-597:331-41. doi: 10.1016/j.scitotenv.2017.04.067 [published Online First: 2017/04/25]

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. *Canadian journal of public health = Revue canadienne de sante publique* 2014;105(1):e69-78. [published Online First: 2014/04/17]

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 promotion and chronic disease prevention in Canada : research, policy and practice* 2016;36(4):76-86. [published Online First: 2016/04/15]

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 : CCC* 2013;24(11):2013-20. doi: 10.1007/s10552-013-0278-x [published Online First: 2013/08/29]

9. Brenner DR, Grevers X, Grundy A, et al. Cancer incidence attributable to inadequate physical activity in Alberta, Canada in 2012. *CMAJ Open* 2016;Under review

10. Brenner DR, Grundy A, Poirier AE, et al. Canver Incidence Attributable to Excess Body Weight in Alberta, Canada. *CMAJ Open* 2016;Submitted Manuscript

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(4):E754-e59. doi: 10.9778/cmajo.20160046 [published Online First: 2016/12/27]

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(4):E768-e75. doi: 10.9778/cmajo.20160036 [published Online First: 2016/12/27]

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(4):E760-e67. doi: 10.9778/cmajo.20160037 [published Online First: 2016/12/27]

14. Grundy A, Poirier AE, Khandwala F, et al. Cancer incidence attributable to insufficient fibre consumption in Alberta in 2012. *CMAJ Open* 2017;5(1):E7-E13. doi: 10.9778/cmajo.20160043

15. Poirier AE, Grundy A, Khandwala F, et al. Cancer incidence attributable to tobacco in Alberta, Canada, in 2012. *CMAJ Open* 2016;4(4):E578-e87. doi: 10.9778/cmajo.20150069 [published Online First: 2016/12/27]

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(3):e0149669. doi: 10.1371/journal.pone.0149669

17. Baade PD, Meng X, Sinclair C, et al. Estimating the future burden of cancers preventable by better diet and physical activity in Australia. *The Medical journal of Australia* 2012;196(5):337-40. [published Online First: 2012/03/22]

18. Pandeya N, Wilson LF, Webb PM, et al. Cancers in Australia in 2010 attributable to the consumption of alcohol. *Australian and New Zealand journal of public health* 2015;39(5):408-13. doi: 10.1111/1753-6405.12456

19. Kendall BJ, Wilson LF, Olsen CM, et al. Cancers in Australia in 2010 attributable to overweight and obesity. *Australian and New Zealand journal of public health* 2015;39(5):452-7. doi: 10.1111/1753-6405.12458 [published Online First: 2015/10/07]

20. Nagle CM, Wilson LF, Hughes MCB, et al. Cancers in Australia in 2010 attributable to the consumption of red and processed meat. *Australian and New Zealand journal of public health* 2015;39(5):429-33. doi: 10.1111/1753-6405.12450

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. *The Lancet Oncology* 2015;16(1):36-46. doi: 10.1016/S1470-2045(14)71123-4

22. Whiteman DC, Webb PM, Green AC, et al. Cancers in Australia in 2010 attributable to modifiable factors: introduction and overview. *Australian and New Zealand journal of public health* 2015;39(5):403-07. doi: 10.1111/1753-6405.12468

23. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3(11):e442. doi: 10.1371/journal.pmed.0030442 [published Online First: 2006/11/30]

24. World Health Organization. Global health risks: mortality and burden of disease attributable to selected major risks: World Health Organization 2009.

25. Parkin DM. 1. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *British journal of 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(3):209-12. doi: 10.1136/jech.2009.090274 [published Online First: 2009/08/21]

27. Levin ML. The occurrence of lung cancer in man. *Acta - Unio Internationalis Contra Cancrum* 1953;9(3):531-41. [published Online First: 1953/01/01]

28. Morgan SL, Winship C. Counterfactuals and causal inference: Cambridge University Press 2014.

29. 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. *Journal of community health* 1982;7(4):292-309. [published Online First: 1982/01/01]

30. 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.

31. International Agency for Reseach on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Lyon, France2017 [Available from: http://monographs.iarc.fr/ accessed June 29 2017.

32. Statistics Canada. Canadian Cancer Registry (CCR) Ottawa, ON: Statistics Canada; 2014 [Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=329031 accessed April 27 2017.

33. Greenland S. Interval estimation by simulation as an alternative to and extension of confidence intervals. *Int J Epidemiol* 2004;33(6):1389-97. doi: 10.1093/ije/dyh276 [published Online First: 2004/08/21]

34. 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. *The Lancet* 2012;380(9838):219-29. doi: 10.1016/S0140-6736(12)61031-9

35. Renehan AG, Soerjomataram I, Tyson M, et al. Incident cancer burden attributable to excess body mass index in 30 European countries. *International journal of cancer Journal international du cancer* 2010;126(3):692-702. doi: 10.1002/ijc.24803 [published Online First: 2009/08/01]

36. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: age-period-cohort models. *Stat Med* 1987;6:469-81.

37. Moller B. Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Cancer Prev* 2002;11(Suppl.):S1-S96.

38. 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. *Technique Report for Cancer Projections Network (C-Proj) Alberta: Alberta Health Services* 2011

39. Dyba T, Hakulinen T, Paivarinta L. A simple non-linear model in incidence prediction. *Statist Med* 1997;16:2297-309.

40. Carstensen B. Age-period-cohort models for the Lexis diagram. *Statistics in medicine* 2007;26(15):3018-45. doi: 10.1002/sim.2764 [published Online First: 2006/12/21]

41. Kim HJ, Fay MP, Feuer EJ, et al. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in medicine* 2000;19(3):335-51. [published Online First: 2000/01/29]

42. Bray I. Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2002;51(2):151-64. doi: 10.1111/1467-9876.00260

43. 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(1):789. doi: 10.1186/1471-2458-10-789

44. Samad AKA, Taylor RS, Marshall T, et al. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Disease* 2005;7(3):204-13. doi: 10.1111/j.1463-1318.2005.00747.x

45. Elm Ev, Altman DG, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335(7624):806-08. doi: 10.1136/bmj.39335.541782.AD

46. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Jama* 2000;283(15):2008-12. doi: 10.1001/jama.283.15.2008

List of Tables and Figures with Captions

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 (second-hand smoke) • UVR risk behaviours • Disinfection by-products • Low vitamin D • Low dietary calcium intake • *Helicobacter pylori* • Hepatitis B • Hepatitis C
$Formula\ 2: PAR = P_c$	• Human papillomavirus • Epstein-Barr virus • Human T-cell lymphotropic virus type 1 • Human herpesvirus 8
$Formula\ 3:\ PAR = \frac{(p_{e1} \times ERR_1) + (P_{e2} \times ERR_2) + \ldots + (P_{ex} \times ERR_x)}{1+\left((p_{e1} \times ERR_1) + (P_{e2} \times ERR_2) + \ldots + (P_{ex} \times ERR_x)\right)}$	• 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
Individualized 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, PAR=population attributable risk, Pe=prevalence of exposure in the population, RR=relative risk, UV=ultraviolet; UVR=ultraviolet radiation; P_c= prevalence of exposure among cases

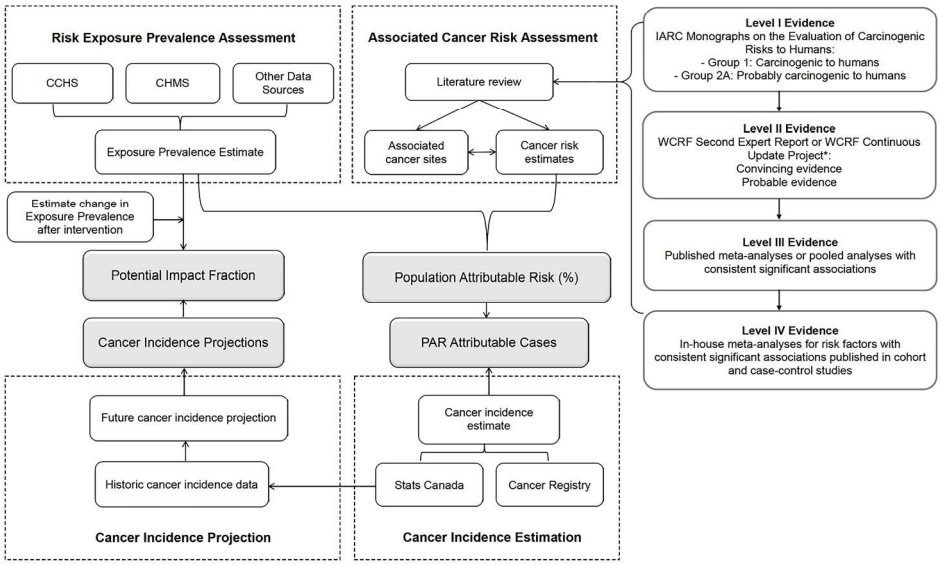


Figure 1. Scope of project framework for estimation of current attributable and future avoidable disease burden. CHMS=Canadian Health Measures Survey, CCHS=Canadian Community Health Survey, Canadian, IARC=International Agency for Research on Cancer, WCRF=World Cancer Research Fund, CUP=Continuous Update Project, PIF=Potential Impact Fraction, PAR Population Attributable Risk

279x215mm (300 x 300 DPI)

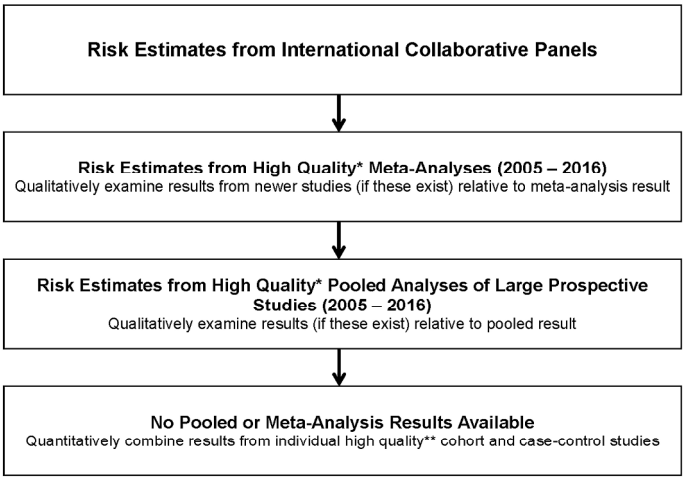


Figure 2. The process flow used for selecting risk estimates used in the ComPARe project.
*Quality determined using STrengthening the Reporting of OBservational studies in Epidemiology (STROBE)[45] guidelines for cohort and case-control studies and Meta-analysis Of Observational Studies in Epidemiology (MOOSE)[46] guidelines for meta-analysis

279x431mm (300 x 300 DPI)

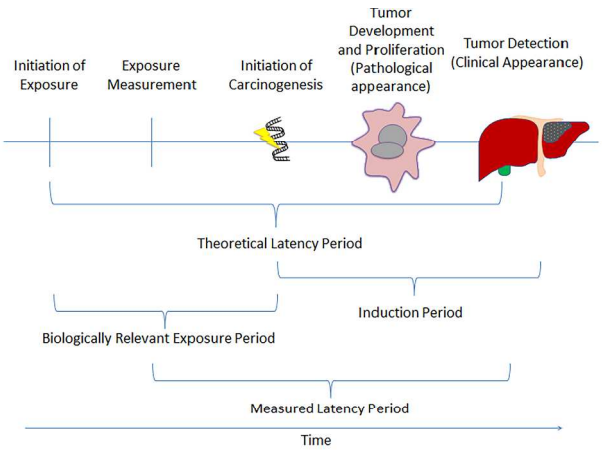
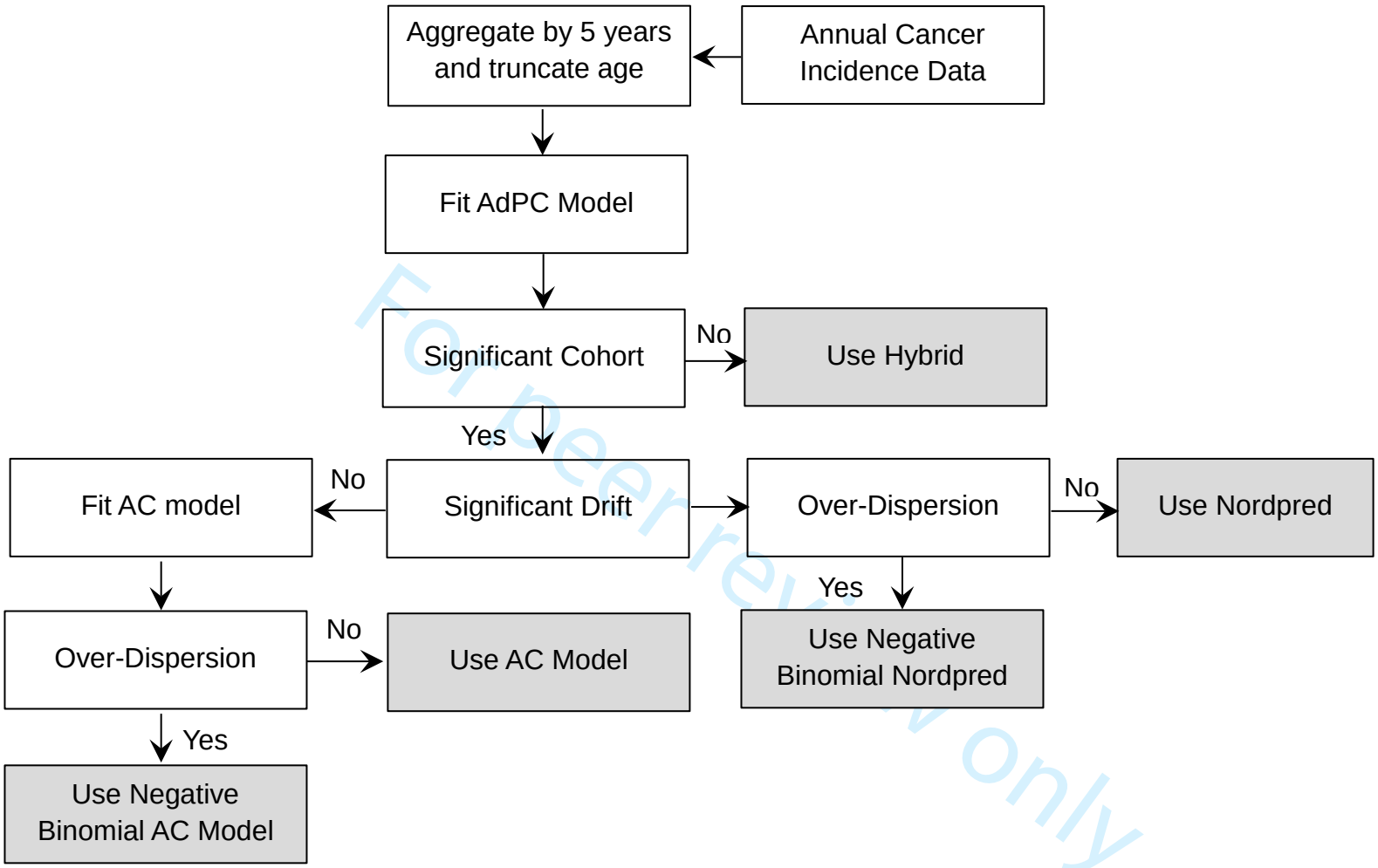


Figure 3. Representation of relevant exposure windows and latency onset considered for the ComPARe project.

279x452mm (300 x 300 DPI)



Supplementary Figure 1. Decision tree for cancer incidence projection model selection in Canproj.AC= Age-cohort model, AdPC= Age-drift-period-cohort model, Hybrid=age-only model or age-period model

*Adapted from: *Canproj-The R package of cancer projection methods based on generalized linear models for age, period and/or cohort. Alberta Health Services: 2011-12-16.*

Supplementary Table 1. Exposure and Cancer Site Associations to be Included in the ComPARe project*

Cancer Site of Interest	Associated Modifiable Risk Factors
Lung	Non-starchy vegetable intake Fruit intake Physical activity Active tobacco smoking Passive tobacco smoking Air pollution Radon Arsenic
Breast	Oral contraceptives Hormone replacement therapy Insufficient fruit Alcohol Red meat Processed meat Insufficient vitamin D Overweight/Obesity Physical inactivity Sedentary behavior Abdominal obesity
Colorectal Cancer	Insufficient fruit Insufficient non-starchy vegetables Alcohol Red meat intake Processed meat intake Fiber intake Insufficient vitamin D Insufficient calcium Overweight/Obesity Physical inactivity Sedentary behavior Abdominal obesity Tobacco smoking
Gastric cancer, Gastric cardia cancer	Insufficient fruit Alcohol Red meat Processed meat Insufficient fibre Overweight/Obesity Tobacco smoking *Helicobacter pylori* (non-cardia only)
Oesophagus Cancer	Overweight/Obesity Physical inactivity Tobacco smoking Alcohol Insufficient non-starchy vegetables Insufficient fruit Processed meat

Bladder Cancer	Insufficient non-starchy vegetables Insufficient fruit Physical inactivity Tobacco smoking Insufficient vitamin D Arsenic Disinfection by-products
Pancreas Cancer	Insufficient non-starchy vegetables Insufficient fruit Alcohol Red meat Processed meat Overweight/Obesity Abdominal obesity Tobacco smoking
Endometrial Cancer	Oral contraceptives Hormone therapy Overweight/Obesity Physical inactivity Sedentary behavior Abdominal obesity
Oral Cancer / Oropharynx Cancer	Insufficient non-starchy vegetables Tobacco smoking Human papillomavirus (HPV)
Liver Cancer	Insufficient non-starchy vegetables Alcohol Overweight/Obesity Physical inactivity Tobacco smoking Hepatitis B virus (HBV) Hepatitis C virus (HCV)
Ovarian Cancer	Oral contraceptives Hormone Replacement therapy Insufficient non-starchy vegetables Overweight/Obesity Sedentary behavior Tobacco smoking
Larynx Cancer	Alcohol Insufficient non-starchy vegetables Tobacco smoking Human papillomavirus (HPV)
Cervical Cancer	Tobacco smoking Passive (second-hand) tobacco smoking Human papillomavirus (HPV)
Prostate Cancer	Overweight/Obesity Abdominal obesity
Kidney Cancer	Overweight/Obesity Abdominal obesity Physical inactivity Tobacco smoking Vitamin D

Gallbladder Cancer	Tobacco smoking Overweight/Obesity
Melanoma	Ultraviolet radiation (indoor tanning, sunburn, sunbathing, total exposure)
Hodgkin Lymphoma	Epstein-Barr Virus (EBV)
Non-Hodgkin Lymphoma	Physical inactivity Epstein-Barr Virus (EBV)(immuno-suppressed only) Hepatitis C virus (HCV)
Non-Melanoma Skin Cancer	Human Immunodeficiency virus (HIV) Ultraviolet radiation (indoor tanning, sunburn, sunbathing, total exposure)
Nasopharyngeal Cancer	Epstein-Barr Virus (EBV)
Pharynx Cancer	HPV
Thyroid Cancer	Overweight/Obesity Abdominal Obesity
Anal Cancer	Human papillomavirus (HPV)
Cholangiocarcinoma	Hepatitis B virus (HBV) Hepatitis C virus (HCV)
Leukemia	Human T-cell lymphotropic virus type 1 (HTLV-1)
Burkitt Lymphoma	Epstein-Barr Virus (EBV)
Eye Cancer	UV radiation
Kaposi Sarcoma	Human herpesvirus 8
Lip Cancer	UV radiation (indoor tanning, sunburn, sunbathing, total exposure)
Extranodal NK/T cell lymphoma – nasal type	Epstein-Barr Virus (EBV)
Mesothelioma	Asbestos
Penile Cancer	Human papillomavirus (HPV)
Tonsil Cancer	Human papillomavirus (HPV)
Vaginal Cancer	Human papillomavirus (HPV)
Vulvar Cancer	Human papillomavirus (HPV)

*inclusion of exposure and cancer site associations were based on hierarchy of evidence collected from the International Agency for Research on Cancer monograph series, World Cancer Research Fund (WCRF) Second Export Report, WCRF Continuous Update Projects and published meta-analyses of epidemiologic studies.

BMJ Open

Estimating the Current and Future Cancer Burden in Canada: Methodologic Framework of the Canadian Population Attributable Risk of Cancer (ComPARe) Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-022378.R1
Article Type:	Protocol
Date Submitted by the Author:	01-Jun-2018
Complete List of Authors:	Brenner, Darren; University of Calgary Cumming School of Medicine, Oncology; Alberta Health Services, Cancer Epidemiology and Prevention Research Poirier, Abbey; Alberta Health Services, Cancer Epidemiology and Prevention Research Walter, Stephen ; McMaster University , Department of Health Research Methods, Evidence, and Impact King, Will; Queen's University, Department of Public Health Sciences Franco, Eduardo; McGill University, Department of Oncology; McGill University Department of Epidemiology Biostatistics and Occupational Health Demers, Paul; Cancer Care Ontario, Occupational Cancer Research Centre Villeneuve, Paul; Carleton University Department of Health Sciences Ruan, Yibing; Alberta Health Services, Cancer Epidemiology and Prevention Research Khandwala, Farah; Alberta Health Services, Cancer Epidemiology and Prevention Research Grevers, Xin; Alberta Health Services, Cancer Epidemiology and Prevention Research Nuttall, Robert; Health Quality Ontario Smith, Leah; Canadian Cancer Society De, Prithwish; Cancer Care Ontario Volesky, Karena; McGill University, Department of Oncology; McGill University Department of Epidemiology Biostatistics and Occupational Health O'Sullivan, Dylan; Queen's University, Department of Public Health Sciences Hystad, Perry ; Oregon State University Friedenreich, CM; Alberta Health Services, Cancer Epidemiology and Prevention Research; University of Calgary Cumming School of Medicine, Oncology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health, Oncology
Keywords:	Epidemiology < ONCOLOGY, Potential impact fraction, Population attributable risk, Cancer, STATISTICS & RESEARCH METHODS, Risk factors

Estimating the Current and Future Cancer Burden in Canada: Methodologic 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 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[12]

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, Montréal, Québec, Canada
6. Department of Oncology, McGill University, Montréal, Québec, Canada
7. Occupational Cancer Research Centre, 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, United States
12. Additional members of the ComPARe study group:
    Elizabeth Holmes[9]
    Zeinab El-Masri[10]
    Mariam El-Zein[6]
    Sheila Bouten[6]
    Tasha Narain[4]
    Priyanka Gogna[4]

*To Whom Correspondence should be addressed:
Darren R. Brenner
Department of Cancer Epidemiology and Prevention Research
CancerControl Alberta, Alberta Health Services
Holy Cross Centre – Room 513C
Box ACB, 2210-2nd St. SW, Calgary, AB
T2S 3C3
Phone: 1-403-698-8178
Darren.Brenner@ucalgary.ca

ABSTRACT

Introduction: The Canadian Population Attributable Risk of Cancer (ComPARe) project will quantify the number and proportion of incident cancer cases 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: Working as 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 for 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 programs 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 cancer are required to allocate 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 aging populations. In addition, despite established associations between modifiable risk factors and cancer risk, sufficient reductions in the prevalence of these risk factors have 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 programs 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 future avoidable national [16-20] and global [21] cancer burdens 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 methodologic extensions to these approaches, including modeling 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 of the impact of potential reductions in exposure prevalence on cancer incidence at 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 standardized 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. Here 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 methodologic framework for the ComPARe study. The concept of PAR or population attributable fraction (PAF) 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 the potential impact fraction (PIF). The PIF is an extension of the PAR to consider situations complete removal of the exposure cannot be assumed.[32] The impact of a reduction in the prevalence or population distribution of an exposure and the subsequent impact of an exposure reduction is examined. The PAR and PIF are 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 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 epidemiologic and basic science literature evaluating etiologic 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 (STROBE)[33] guidelines for cohort and case-control studies and Meta-analysis Of Observational Studies in Epidemiology (MOOSE)[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 epidemiologic 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 Supplementary Table 1.

Estimation of Attributable Cancers

Exposure Prevalence Data – Including Latency

The biologically relevant time period from 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 tumors. Therefore, we allowed for a period of latency from exposure to cancer incidence/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 *H. pylori*, there was one seroprevalence survey in 1999-2000, and for HBV & HCV the prevalence data were collected from the Canadian Health Measures and the Canadian Notifiable Disease Surveillance System occurring from 2007-2012. 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 (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 five-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 (ICD-O-3). 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 modeling 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% confidence intervals (CIs) cannot be easily calculated for PARs[38], Monte Carlo simulation methods were used to estimate 95% CIs around PAR estimates, where the 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 (e.g. 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 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 the 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 /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 (e.g., colorectal, breast, lung, and cervical cancers) or reduced (e.g., 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 (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 of these methods to fit the best model. We evaluated all 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 which 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, e.g., 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/interactions on cancer risk. Several well-characterized examples include alcohol and tobacco for various cancer sites [48], and overweight/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 characterize 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 (i.e. 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 policy makers, clinicians and the public to inform priority setting in prevention programming and resources; allocation of funding to areas of unmet need; etc. We have developed this project in collaboration with our knowledge translation partner – the Canadian Cancer Society (CCS). 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.

Methodologic Extensions

During this project we encountered several methodologic 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, 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 programs.

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/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, *Helicobacter 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 United States 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-cardia 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 program [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 in the future as the models do not account for future changes in risk factors (i.e. 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 Canadian Cancer Registry 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 would 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 methodologic 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 policy makers, healthcare providers and the general public for a wide range of applications in cancer control and prevention.

Competing Interests

We have read and understood the BMJ Group policy on declaration of interests and declare no competing interests.

Contributor Statement

D.R.B, C.M.F., S.D.W., W.D.K, E.L.F., P.D., P.J.V., R.N., and P.De. were responsible for the study conception. D.R.B., A.E.P., S.D.W., W.D.K., E.L.F., P.D., P.J.V., Y.R., F.K., X.G., R.N, L.S., P.De., K.V., D.O., P.H, C.M.F. contributed substantially to the study design. D.R.B, A.E.P. and Y.R. drafted the manuscript. D.R.B., A.E.P., S.D.W., W.D.K., E.L.F., P.D., P.J.V., Y.R., F.K., X.G., R.N, L.S., P.De., K.V., D.O., P.H, C.M.F. revised the draft paper, 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.

Licence to BMJ Publishing Group Limited

I Darren Brenner, The Corresponding Author of this article contained within the original manuscript which includes any diagrams & photographs within and any related or stand-alone film submitted has the right to grant on behalf of all authors and does grant on behalf of all authors, a licence to the BMJ Publishing Group Ltd and its licencees, to permit this Contribution (if accepted) to be published in the BMJ Open and any other BMJ Group products and to exploit all subsidiary rights.

Transparency Declaration

Darren Brenner affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Funding

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

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 promotion and chronic disease prevention in Canada : research, policy and practice* 2015;35 Suppl 1:2-186. [published Online First: 2015/05/27]

3. Bray F, Moller B. Predicting the future burden of cancer. *Nat Rev Cancer* 2006;6(1):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. *Radiation protection dosimetry* 2012;152(1-3):9-13. doi: 10.1093/rpd/ncs147 [published Online First: 2012/08/10]

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. *The Science of the total environment* 2017;596-597:331-41. doi: 10.1016/j.scitotenv.2017.04.067 [published Online First: 2017/04/25]

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. *Canadian journal of public health = Revue canadienne de sante publique* 2014;105(1):e69-78. [published Online First: 2014/04/17]

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 promotion and chronic disease prevention in Canada : research, policy and practice* 2016;36(4):76-86. [published Online First: 2016/04/15]

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 : CCC* 2013;24(11):2013-20. doi: 10.1007/s10552-013-0278-x [published Online First: 2013/08/29]

9. Brenner DR, Grevers X, Grundy A, et al. Cancer incidence attributable to inadequate physical activity in Alberta, Canada in 2012. *CMAJ Open* 2016;Under review

10. Brenner DR, Grundy A, Poirier AE, et al. Canver Incidence Attributable to Excess Body Weight in Alberta, Canada. *CMAJ Open* 2016;Submitted Manuscript

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(4):E754-e59. doi: 10.9778/cmajo.20160046 [published Online First: 2016/12/27]

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(4):E768-e75. doi: 10.9778/cmajo.20160036 [published Online First: 2016/12/27]

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(4):E760-e67. doi: 10.9778/cmajo.20160037 [published Online First: 2016/12/27]

14. Grundy A, Poirier AE, Khandwala F, et al. Cancer incidence attributable to insufficient fibre consumption in Alberta in 2012. *CMAJ Open* 2017;5(1):E7-E13. doi: 10.9778/cmajo.20160043

15. Poirier AE, Grundy A, Khandwala F, et al. Cancer incidence attributable to tobacco in Alberta, Canada, in 2012. *CMAJ Open* 2016;4(4):E578-e87. doi: 10.9778/cmajo.20150069 [published Online First: 2016/12/27]

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(3):e0149669. doi: 10.1371/journal.pone.0149669

17. Baade PD, Meng X, Sinclair C, et al. Estimating the future burden of cancers preventable by better diet and physical activity in Australia. *The Medical journal of Australia* 2012;196(5):337-40. [published Online First: 2012/03/22]

18. Pandeya N, Wilson LF, Webb PM, et al. Cancers in Australia in 2010 attributable to the consumption of alcohol. *Australian and New Zealand journal of public health* 2015;39(5):408-13. doi: 10.1111/1753-6405.12456

19. Kendall BJ, Wilson LF, Olsen CM, et al. Cancers in Australia in 2010 attributable to overweight and obesity. *Australian and New Zealand journal of public health* 2015;39(5):452-7. doi: 10.1111/1753-6405.12458 [published Online First: 2015/10/07]

20. Nagle CM, Wilson LF, Hughes MCB, et al. Cancers in Australia in 2010 attributable to the consumption of red and processed meat. *Australian and New Zealand journal of public health* 2015;39(5):429-33. doi: 10.1111/1753-6405.12450

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. *The Lancet Oncology* 2015;16(1):36-46. doi: 10.1016/S1470-2045(14)71123-4

22. Whiteman DC, Webb PM, Green AC, et al. Cancers in Australia in 2010 attributable to modifiable factors: introduction and overview. *Australian and New Zealand journal of public health* 2015;39(5):403-07. doi: 10.1111/1753-6405.12468

23. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3(11):e442. doi: 10.1371/journal.pmed.0030442 [published Online First: 2006/11/30]

24. World Health Organization. Global health risks: mortality and burden of disease attributable to selected major risks: World Health Organization 2009.

25. Parkin DM. 1. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *British journal of cancer* 2011;105(S2):S2-S5.

26. Barendregt JJ, Veerman JL. Categorical versus continuous risk factors and the calculation of potential impact fractions. *Journal of epidemiology and community health* 2010;64(3):209-12. doi: 10.1136/jech.2009.090274 [published Online First: 2009/08/21]

27. Arnold M, Touillaud M, Dossus L, et al. Cancers in France in 2015 attributable to high body mass index. *Cancer epidemiology* 2018;52:15-19. doi: 10.1016/j.canep.2017.11.006 [published Online First: 2017/11/22]

28. Gwenn M, Ivana K, Joséphine B, et al. Tobacco-attributable burden of cancer according to socioeconomic position in France. *International Journal of Cancer*;0(0) doi: doi:10.1002/ijc.31328

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 (Abingdon, England)* 2018;113(2):247-56. doi: 10.1111/add.14009 [published Online First: 2017/08/24]

30. Levin ML. The occurrence of lung cancer in man. *Acta - Unio Internationalis Contra Cancrum* 1953;9(3):531-41. [published Online First: 1953/01/01]

31. Morgan SL, Winship C. Counterfactuals and causal inference: 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. *Journal of community health* 1982;7(4):292-309. [published Online First: 1982/01/01]

33. Elm Ev, Altman DG, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335(7624):806-08. doi: 10.1136/bmj.39335.541782.AD

34. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Jama* 2000;283(15):2008-12. doi: 10.1001/jama.283.15.2008

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, France2017 [Available from: http://monographs.iarc.fr/ accessed June 29 2017.

37. Statistics Canada. Canadian Cancer Registry (CCR) Ottawa, ON: Statistics Canada; 2014 [Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=329031 accessed April 27 2017.

38. Greenland S. Interval estimation by simulation as an alternative to and extension of confidence intervals. *International journal of epidemiology* 2004;33(6):1389-97. doi: 10.1093/ije/dyh276 [published Online First: 2004/08/21]

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. *The Lancet* 2012;380(9838):219-29. doi: 10.1016/S0140-6736(12)61031-9

40. Renehan AG, Soerjomataram I, Tyson M, et al. Incident cancer burden attributable to excess body mass index in 30 European countries. *International journal of cancer Journal international du cancer* 2010;126(3):692-702. doi: 10.1002/ijc.24803 [published Online First: 2009/08/01]

41. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: age-period-cohort models. *Stat Med* 1987;6:469-81.

42. Moller B. Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Cancer Prev* 2002;11(Suppl.):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. *Technique Report for Cancer Projections Network (C-Proj) Alberta: Alberta Health Services* 2011

44. Dyba T, Hakulinen T, Paivarinta L. A simple non-linear model in incidence prediction. *Statist Med* 1997;16:2297-309.

45. Carstensen B. Age-period-cohort models for the Lexis diagram. *Statistics in medicine* 2007;26(15):3018-45. doi: 10.1002/sim.2764 [published Online First: 2006/12/21]

46. Kim HJ, Fay MP, Feuer EJ, et al. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in medicine* 2000;19(3):335-51. [published Online First: 2000/01/29]

47. Bray I. Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2002;51(2):151-64. doi: 10.1111/1467-9876.00260

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(1):789. doi: 10.1186/1471-2458-10-789

49. Samad AKA, Taylor RS, Marshall T, et al. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Disease* 2005;7(3):204-13. doi: 10.1111/j.1463-1318.2005.00747.x

50. Plummer M, Franceschi S, Vignat J, et al. Global burden of gastric cancer attributable to Helicobacter pylori. *International journal of cancer Journal international du cancer* 2015;136(2):487-90. doi: 10.1002/ijc.28999 [published Online First: 2014/06/04]

List of Tables and Figures with Captions

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 (second-hand 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 herpesvirus 8
$Formula\ 3:\ PAR = \frac{(p_{e1} \times ERR_1) + (P_{e2} \times ERR_2) + \ldots + (P_{ex} \times ERR_x)}{1 + \left((p_{e1} \times ERR_1) + (P_{e2} \times ERR_2) + \ldots + (P_{ex} \times ERR_x)\right)}$	• 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
Individualized 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

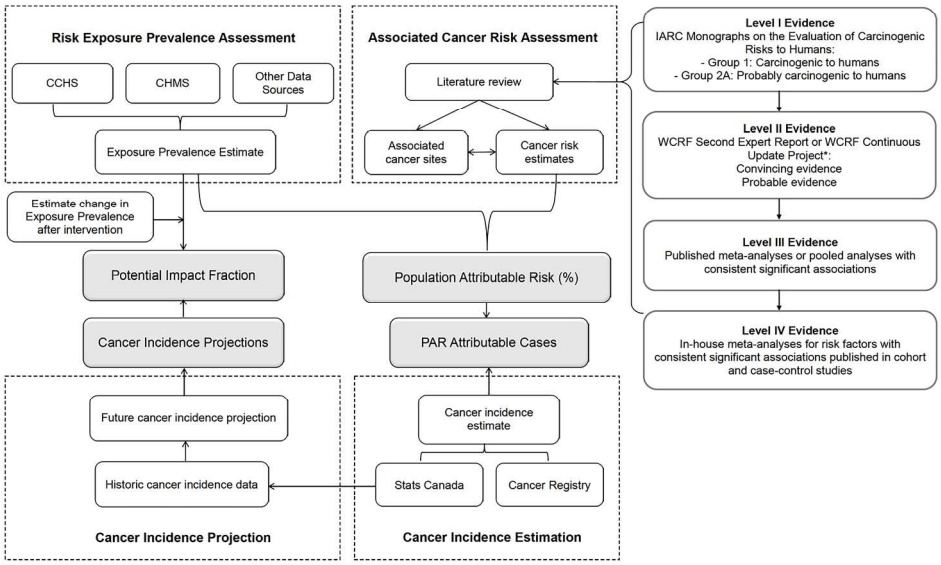


Figure 1. Scope of project framework for estimation of current attributable and future avoidable disease burden. CHMS=Canadian Health Measures Survey, CCHS=Canadian Community Health Survey, Canadian, IARC=International Agency for Research on Cancer, WCRF=World Cancer Research Fund, CUP=Continuous Update Project, PIF=Potential Impact Fraction, PAR Population Attributable Risk

279x215mm (300 x 300 DPI)

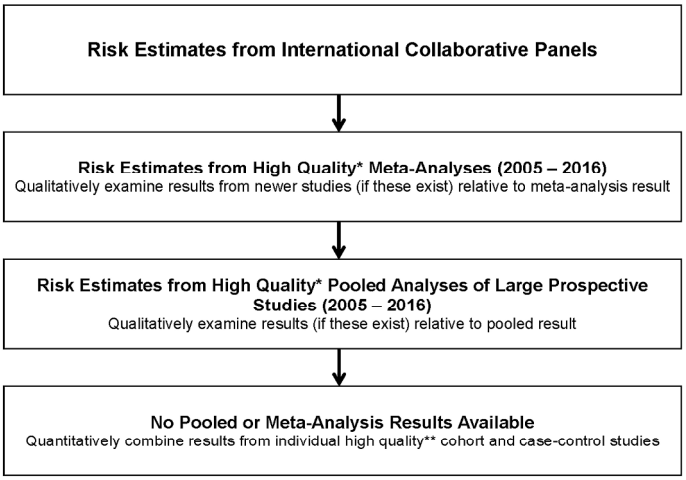


Figure 2. The process flow used for selecting risk estimates used in the ComPARe project.† *Quality determined using STrengthening the Reporting of OBservational studies in Epidemiology (STROBE)[33] guidelines for cohort and case-control studies and Meta-analysis Of Observational Studies in Epidemiology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

279x431mm (300 x 300 DPI)

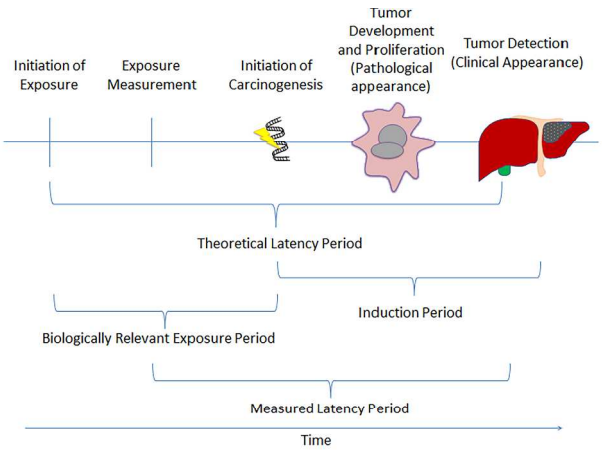


Figure 3. Representation of relevant exposure windows and latency onset considered for the ComPARe project.

279x452mm (300 x 300 DPI)

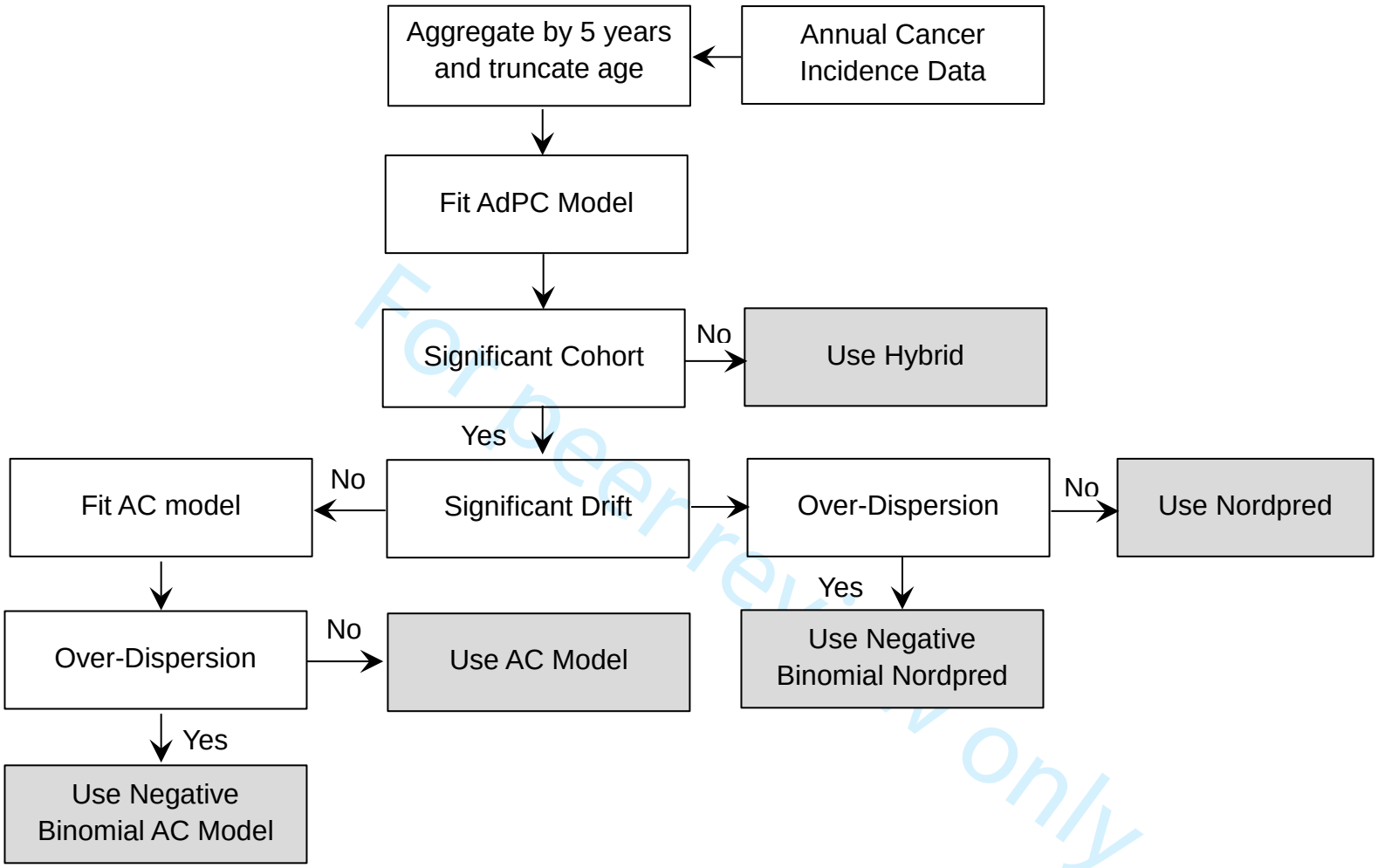
Supplementary Table 1. Exposure and Cancer Site Associations to be Included in the ComPARe project*

Cancer Site of Interest	Associated Modifiable Risk Factors
Lung	Non-starchy vegetable intake Fruit intake Physical activity Active tobacco smoking Passive tobacco smoking Air pollution Radon Arsenic
Breast	Oral contraceptives Hormone replacement therapy Insufficient fruit Alcohol Red meat Processed meat Insufficient vitamin D Overweight/Obesity Physical inactivity Sedentary behavior Abdominal obesity
Colorectal Cancer	Insufficient fruit Insufficient non-starchy vegetables Alcohol Red meat intake Processed meat intake Fiber intake Insufficient vitamin D Insufficient calcium Overweight/Obesity Physical inactivity Sedentary behavior Abdominal obesity Tobacco smoking
Gastric cancer, Gastric cardia cancer	Insufficient fruit Alcohol Red meat Processed meat Insufficient fibre Overweight/Obesity Tobacco smoking *Helicobacter pylori* (non-cardia only)
Oesophagus Cancer	Overweight/Obesity Physical inactivity Tobacco smoking Alcohol Insufficient non-starchy vegetables Insufficient fruit Processed meat

Bladder Cancer	Insufficient non-starchy vegetables Insufficient fruit Physical inactivity Tobacco smoking Insufficient vitamin D Arsenic Disinfection by-products
Pancreas Cancer	Insufficient non-starchy vegetables Insufficient fruit Alcohol Red meat Processed meat Overweight/Obesity Abdominal obesity Tobacco smoking
Endometrial Cancer	Oral contraceptives Hormone therapy Overweight/Obesity Physical inactivity Sedentary behavior Abdominal obesity
Oral Cancer / Oropharynx Cancer	Insufficient non-starchy vegetables Tobacco smoking Human papillomavirus (HPV)
Liver Cancer	Insufficient non-starchy vegetables Alcohol Overweight/Obesity Physical inactivity Tobacco smoking Hepatitis B virus (HBV) Hepatitis C virus (HCV)
Ovarian Cancer	Oral contraceptives Hormone Replacement therapy Insufficient non-starchy vegetables Overweight/Obesity Sedentary behavior Tobacco smoking
Larynx Cancer	Alcohol Insufficient non-starchy vegetables Tobacco smoking Human papillomavirus (HPV)
Cervical Cancer	Tobacco smoking Passive (second-hand) tobacco smoking Human papillomavirus (HPV)
Prostate Cancer	Overweight/Obesity Abdominal obesity
Kidney Cancer	Overweight/Obesity Abdominal obesity Physical inactivity Tobacco smoking Vitamin D

Gallbladder Cancer	Tobacco smoking Overweight/Obesity
Melanoma	Ultraviolet radiation (indoor tanning, sunburn, sunbathing, total exposure)
Hodgkin Lymphoma	Epstein-Barr Virus (EBV)
Non-Hodgkin Lymphoma	Physical inactivity Epstein-Barr Virus (EBV)(immuno-suppressed only) Hepatitis C virus (HCV)
Non-Melanoma Skin Cancer	Human Immunodeficiency virus (HIV) Ultraviolet radiation (indoor tanning, sunburn, sunbathing, total exposure)
Nasopharyngeal Cancer	Epstein-Barr Virus (EBV)
Pharynx Cancer	HPV
Thyroid Cancer	Overweight/Obesity Abdominal Obesity
Anal Cancer	Human papillomavirus (HPV)
Cholangiocarcinoma	Hepatitis B virus (HBV) Hepatitis C virus (HCV)
Leukemia	Human T-cell lymphotropic virus type 1 (HTLV-1)
Burkitt Lymphoma	Epstein-Barr Virus (EBV)
Eye Cancer	UV radiation
Kaposi Sarcoma	Human herpesvirus 8
Lip Cancer	UV radiation (indoor tanning, sunburn, sunbathing, total exposure)
Extranodal NK/T cell lymphoma – nasal type	Epstein-Barr Virus (EBV)
Mesothelioma	Asbestos
Penile Cancer	Human papillomavirus (HPV)
Tonsil Cancer	Human papillomavirus (HPV)
Vaginal Cancer	Human papillomavirus (HPV)
Vulvar Cancer	Human papillomavirus (HPV)

*inclusion of exposure and cancer site associations were based on hierarchy of evidence collected from the International Agency for Research on Cancer monograph series, World Cancer Research Fund (WCRF) Second Export Report, WCRF Continuous Update Projects and published meta-analyses of epidemiologic studies.



Supplementary Figure 1. Decision tree for cancer incidence projection model selection in Canproj.AC= Age-cohort model, AdPC= Age-drift-period-cohort model, Hybrid=age-only model or age-period model

*Adapted from: *Canproj-The R package of cancer projection methods based on generalized linear models for age, period and/or cohort. Alberta Health Services: 2011-12-16.*