PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Estimating the Current and Future Cancer Burden in Canada: Methodologic Framework of the Canadian Population Attributable Risk of Cancer (ComPARe) Study
AUTHORS	Brenner, Darren; Poirier, Abbey; Walter, Stephen; King, Will; Franco, Eduardo; Demers, Paul; Villeneuve, Paul; Ruan, Yibing; Khandwala, Farah; Grevers, Xin; Nuttall, Robert; Smith, Leah; De, Prithwish; Volesky, Karena; O'Sullivan, Dylan; Hystad, Perry; Friedenreich, CM

VERSION 1 – REVIEW

REVIEWER	Manami Inoue National Cancer Center, Japan
REVIEW RETURNED	19-Mar-2018

GENERAL COMMENTS	This is a profile paper of ComPARe study, which is important and crucial for the national level cancer control policy. The manuscript is well-written in general, but there are some minor points to be concerned to improve the worth of this profile paper to be cited by many papers. 1. Latency period (p. 8-) The authors do not use the fixed latency period such as “10” years, but rather select from the studies when available from cohort studies. It is helpful if the authors add some values of distribution of latency periods. If it is more or less 10 years, then authors can use 10 years for all estimates, Previous studies do this way. The authors better provide the reason for not using fixed value, (I understand some exception exists.) 2. Risk factors (p. 7) The authors provided the list of risk factors by each cancer (supplement table 1), but it is helpful if the authors provide the list of cancers by risk factors. 3. Incidence The authors focused incidence only. Justification needed why you did not focus cancer death. 4. Methods of estimating PAR Many previous studies use different method of estimating PAR for infectious agents. The authors should add description to clarify. It is sometimes difficult to estimate by typical methods due to lack of prevalence of specific infectious agents and accordingly uses the PAF value itself to estimate attributable cases. 5. Multiple risk factor method (Ruan 2017) (p. 12) This should be published? Please cite Ruan paper. If not published,

	you should add description of this method in this manuscript. 6. Canada Population Canada, I understand, is multiethnic country. Considering the difference in the prevalence or cancer incidence by ethnic groups and many population in Canada are the first generation, I feel that the authors need to consider somehow the prevalence by ethnic groups and apply the difference into estimation. How does the authors justify this? (For example, agents such as HCV/HBV and H. pylori, they are influenced by mother countries and generation (birth cohort) They also use the relative risk from IARC or WCRF estimates. These agencies published summary relative risks by region such as Asia and Europe. I think these values can be also applied for more precise estimates.

REVIEWER	Martyn Plummer International Agency for Research on Cancer, France
REVIEW RETURNED	03-Apr-2018

GENERAL COMMENTS	Exercises in estimating attributable fractions are becoming increasingly popular. I know from my own experiences in this field (e.g. https://doi.org/10.1016/S2214-109X(16)30143-7) that the key analysis choices are difficult to convey in the final report. So a protocol article that sets out these choices in advance, and provides a methodological framework for the analysis, is a useful contribution. There is a danger that by making some of the difficult choices explicit, the authors open themselves up to criticism. For example, the sentence "Models were selected based on expert opinion of the visual evaluation of the fit to past data trends" does not inspire much confidence. Even when there is a well defined decision tree, such as in supplementary figure 1, this raises its own problems. It is quite likely that the final model chosen by this tree will be over-fitted and will therefore considerably underestimate the uncertainty in the projections. Full accounting of the uncertainty in attributable risk calculations remains an unsolved problem. I am a little concerned that the Canproj package for R is not available from the Comprehensive R Archive Network (https://cran.r-project.org). This means it has not passed the quality assurance tests required by the CRAN maintainers. Furthermore, the reference for Canproj is a technical report from 2011. If this is an important piece of infrastructure for the project then I would recommend at least uploading to CRAN and, if possible, writing an article for peer review in an appropriate journal (e.g. JSS, The R Journal, ...). Minor points A similar project on estimating attributable risks in France has recently been completed and may be worth citing, e.g. https://doi.org/10.1016/j.canep.2017.11.006 https://doi.org/10.1007/s10654-017-0334-z https://doi.org/10.1002/ijc.31328 https://doi.org/10.1007/s10552-018-1015-2 P6. Attributable fractions can also be calculated from the prevalence in cases using Bruzzi's formula. This the approach we have used to estimate AFs for infections. Formula 2 is the limiting case of the Bruzzi formula when the relative risk goes to infinity.

	P7. Does the list of carcinogens identified by IARC include only group 1 ("carcinogenic to humans") or also group 2A ("probably carcinogenic to humans") ? P16 Our latest estimates for the fraction of non-cardia gastric cancer attributable to H. pylori suggests the AF is 89% (https://doi.org/10.1002/ijc.28999). This is based on cohort data from low-risk countries using immunoblot to measure infection.

VERSION 1 – AUTHOR RESPONSE

Reviewer Response – BMJ Open Manuscript ID bmjopen-2018-022378

We thank the reviewers for their detailed and thoughtful reviews. We have revised the manuscript to address their comments and have included a point-by-point response below.

Reviewer: 1

Reviewer Name: Manami Inoue

Institution and Country: National Cancer Center, Japan Competing Interests: None

This is a profile paper of ComPARe study, which is important and crucial for the national level cancer control policy.

The manuscript is well-written in general, but there are some minor points to be concerned to improve the worth of this profile paper to be cited by many papers.

Comment: Latency period (p. 8-)

The authors do not use the fixed latency period such as "10" years, but rather select from the studies when available from cohort studies. It is helpful if the authors add some values of distribution of latency periods. If it is more or less 10 years, then authors can use 10 years for all estimates, Previous studies do this way. The authors better provide the reason for not using fixed value, (I understand some exception exists.)

Response: We have revised our description of the process for selecting latency periods for the array of included exposures in this set of analyses. Flexibility in the latency period is required due to the variable biology of how latency occurs for different cancer sites.

We have now revised the text to reflect the balance between selecting a biologically plausible and relevant period of time as well as the pragmatic nature of prevalence data collection.

For example, for the infectious agents, the latency period was determined by the *availability* of prevalence data. For *H. pylori*, there was one sero-survey in 1999-2000, and for HBV & HCV the

prevalence data come from the Canadian Health Measures and the Canadian Notifiable Disease Surveillance System occurring from 2007-2012.

The text now included the following on page 16 “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 collecte from the Canadian Health Measures and the Canadian Notifiable Disease Surveillance System occurring from 2007-2012.”

Comment: Risk factors (p. 7)

The authors provided the list of risk factors by each cancer (supplement table 1), but it is helpful if the authors provide the list of cancers by risk factors.

Response: The modifiable risk factors for each cancer site were included in the analysis based on various levels of evidence as described in the manuscript. For certain exposures, there is sometimes more uncertainty with one cancer type versus another. Our listing is based on the quantity and strength of evidence, which is curated by cancer site. To list cancers by risk factors implies that the evidence base for a given exposure is comparable across all cancers. For these reasons we have retained the organization of risk factors by cancer site.

Comment: Incidence

The authors focused incidence only. Justification needed why you did not focus cancer death.

Response: In the ComPARe project we are focused on generating data to support cancer prevention initiatives in Canada. For this reason we focused our analyses on modeling cancer incidence. From a pragmatic perspective, including cancer deaths involves making a number of assumption about: changes in treatment, changes in survival following diagnosis (which may or may not be differential based on whether the behaviours (risk factors) are changed) as well as differential survival across risk factor exposure group, where the evidence for effects have not been adequately studied.

The text now includes the following on Page 9: “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.”

Comment: Methods of estimating PAR

Many previous studies use different method of estimating PAR for infectious agents. The authors should add description to clarify. It is sometimes difficult to estimate by typical

methods due to lack of prevalence of specific infectious agents and accordingly uses the PAF value itself to estimate attributable cases.

We thank the reviewer for the comment. The methods used in other studies estimating PARs for infections (Shin 2011; Parkin 2011; de Martel 2012; Antonsson 2015; Plummer 2016; Silva 2016) informed our own methods. Our methods are in line with what has been performed previously, with the exception that we do not utilize the AR equation that uses prevalence in cases to substitute prevalence in the population because that equation requires additional assumptions and population prevalence estimates were available, therefore we use the preferred equation (the one with population prevalence). For infection-cancer sites pairs where the relative risk is very high (tends to infinity, as is case for EBV, HPV, HHV-8 & HTLV-1) we use the prevalence in cases to approximate the PAF.

Comment: Multiple risk factor method (Ruan 2017) (p. 12)

This should be published? Please cite Ruan paper. If not published, you should add description of this method in this manuscript.

Response: Given that this work from our team is not yet published we have revised this section and included a description of the Miettinen-Steenland. The Miettinen-Steenland method is practical approach.

In the text: We have removed the discussion of the Ruan paper (under consideration) and included the following on page 13: ”In order to combine PAR across exposures we used the Miettinen-Steenland Approach for any combined or “summary” estimates.”

Comment: Canada Population

Canada, I understand, is multiethnic country. Considering the difference in the prevalence for cancer incidence by ethnic groups and many population in Canada are the first generation, I feel that the authors need to consider somehow the prevalence by ethnic groups and apply the difference into estimation. How does the authors justify this? (For example, agents such as HCV/HBV and H. pylori, they are influenced by mother countries and generation (birth cohort) They also use the relative risk from IARC or WCRF estimates. These agencies published summary relative risks by region such as Asia and Europe. I think these values can be also applied for more precise estimates.

The limitation of not including ethnicity in the estimation has now been addressed in the discussion. In order to consider ethnicity in our estimates, we would require prevalence, risk estimates (assuming associations are modified by ethnicity) and cancer incidence data by ethnicity, which are not available for most exposures. In addition, characterizing lifetime exposure in immigrants would be extremely challenging as residential history data are not available. For some infections, country of origin is relevant (*H. pylori* and HBV). One of the points which will be discussed in an the infection specific PAR manuscript is the level of coverage of the population-based prevalence estimates and potential

limitations of the sources. For ultraviolet radiation (where there is a strong interaction by ethnicity) we have actually taken ethnicity into account, which is an improvement on previous PAR studies for UVR.

We have included the following clarification for ethnicity on Page 16: “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.”

Reviewer: 2

Reviewer Name: Martyn Plummer

Institution and Country: International Agency for Research on Cancer, France Competing Interests: None declared

Exercises in estimating attributable fractions are becoming increasingly popular. I know from my own experiences in this field (e.g. https://doi.org/10.1016/S2214-109X(16)30143-7) that the key analysis choices are difficult to convey in the final report. So a protocol article that sets out these choices in advance, and provides a methodological framework for the analysis, is a useful contribution.

Comment: There is a danger that by making some of the difficult choices explicit, the authors open themselves up to criticism. For example, the sentence "Models were selected based on expert opinion of the visual evaluation of the fit to past data trends" does not inspire much confidence. Even when there is a well defined decision tree, such as in supplementary figure 1, this raises its own problems. It is quite likely that the final model chosen by this tree will be over-fitted and will therefore considerably underestimate the uncertainty in the projections. Full accounting of the uncertainty in attributable risk calculations remains an unsolved problem.

Response: We thank the reviewer for their comment. We have revised our discussion to further address limitations in selecting cancer incidence projection models and to highlight that for several of the issues of interest - there is no correct answer. We feel that it is helpful to a broad readership base

to explicitly outline our choices for modeling. It can be argued that our approach is more transparent as it highlights how we dealt with what we believe are the sources of uncertainty.

Comment: I am a little concerned that the Canproj package for R is not available from the Comprehensive R Archive Network (https://cran.r-project.org). This means it has not passed the quality assurance tests required by the CRAN maintainers. Furthermore, the reference for Canproj is a technical report from 2011. If this is an important piece of infrastructure for the project then I would recommend at least uploading to CRAN and, if possible, writing an article for peer review in an appropriate journal (e.g. JSS, The R Journal, ...).

Response: We agree with the reviewer that it is unfortunate that the Canproj package for R is not available from the R archive network. However, the Canproj package simply incorporates a decision tree to choose the best projection model from widely accepted models such as NordPred and negative binomials, as described in the manuscript (page 11). In addition, as stated in the methods of the manuscript, all projections were evaluated, independently of goodness-of-fit, to inspect the face validity of the projections. We are working with the individual who developed the Canproj package to have it uploaded to CRAN.

Minor points

Comment: A similar project on estimating attributable risks in France has recently been completed and may be worth citing, e.g.

https://doi.org/10.1016/j.canep.2017.11.006

https://doi.org/10.1007/s10654-017-0334-z

https://doi.org/10.1002/ijc.31328

https://doi.org/10.1007/s10552-018-1015-2

Response: We have included the citations for the work completed in France. We thank the reviewer for highlighting this work.

Comment: Attributable fractions can also be calculated from the prevalence in cases using Bruzzi's formula. This the approach we have used to estimate AFs for infections. Formula 2 is the limiting case of the Bruzzi formula when the relative risk goes to infinity.

Response: We have revised Formula 2 with the formula from Bruzzi. We feel that this aligns with our analytical approach as the reviewer has highlighted that Formula 2 as previously presented is the limiting case of the Bruzzi formula.

We feel that this change in fact best reflects our approach as for infection-cancer site pairs where the infection is a necessary cause (e.g. HPV in cervical cancer, HTLV-1 in adult T-cell leukemia/lymphoma, and HHV-8 in Kaposi sarcoma), there are no questions about using prevalence in cases. There are many more sites associated with EBV and HPV where we use the prevalence in cases to approximate PAR using the Bruzzie formula.

Comment: Does the list of carcinogens identified by IARC include only group 1 ("carcinogenic to humans") or also group 2A ("probably carcinogenic to humans") ?

Response: The list of carcinogens includes both group 1 and 2A. This has been clarified in the manuscript.

Comment: Our latest estimates for the fraction of non-cardia gastric cancer attributable to H. pylori suggests the AF is 89% (https://doi.org/10.1002/ijc.28999). This is based on cohort data from low-risk countries using immunoblot to measure infection.

Response: We have included the following revision to the manuscript:

"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. (Plummer 2016). To account for the new gold standard, the included studies will be corrected for measurement error."