ComPARe statistical methods – An overview

Contents

Introduction

The Canadian Population Attributable Risk of Cancer (ComPARe) study estimated the current and future burden of cancer caused by modifiable risk factors.

Population attributable risk (PAR) is the measure used to estimate both the current and future burden of cancer attributed to a risk factor for cancer. A statistical measure called the potential impact fraction (PIF) was also used to estimate the future avoidable burden of cancer if exposure to a risk factor was reduced.

A detailed description of the methodological framework used for this study was previously published in *BMJ Open*. The current document provides a summary of notable aspects of the methods. A **Glossary** to define terms used in ComPARe is also provided at the end of this document.

How risk factors and cancers were selected

PARs and PIFs are based on the assumption that there is a cause and effect relationship between exposure to a risk factor and developing cancer. Therefore, only exposure-cancer pairs with strong supporting evidence of an association were considered.

An exposure-cancer pair was eligible for inclusion if:

- the International Agency for Research on Cancer (IARC) classified the exposure as either *carcinogenic to humans* (*Group 1*) or *probably carcinogenic to humans* (*Group 2A*), or
- the World Cancer Research Fund (WCRF) classified the exposure as having *convincing* or *probable* evidence, or
- meta-analyses or pooled analyses of high-quality epidemiologic studies produced after the WCRF and IARC reports demonstrated consistent and statistically significant associations for an exposure-cancer pair.

A complete list of risk factors examined in ComPARe can be found in Table 1 and a list of cancers examined is shown in Table 2. In some instances, exposure-cancer pairs met the inclusion criteria above but were not included in the study due to data limitations.

A matrix of all the exposure-cancer pairs examined in ComPARe is available.

Grouped risk factors

Sometimes it is useful to know estimates for risk factors that can naturally be grouped together. For example, a summary estimate of attributable risk for “healthy living” requires combining individual risk factors on diet, physical inactivity, sedentary behaviour and weight. Grouped risk factors in ComPARe are listed in Table 1.

Joint effects of 2 or more risk factors

Some exposures are known to interact together to further increase cancer risk compared to their individual impacts. For example, when obesity exists alongside physical inactivity, they have a synergistic (i.e. joint) effect on the risk of colorectal cancer that is different from the effect of each risk factor alone. Estimating the PAR for joint risk factors is challenging because knowledge about their joint prevalence and joint relative risks (RRs) is needed but often unavailable. The PARs for a limited number of such joint risk factors have been estimated (see Table 1).

Occupational risk factors

A separate study on the occupational burden of cancer was led by the Occupational Cancer Research Centre (OCRC) at Cancer Care Ontario. Some of the results of that study are presented alongside those of ComPARe in order to provide a more comprehensive picture of the preventable burden of cancer in Canada. Those occupational exposures are described in Table 1.

How the current attributable burden of cancer was estimated

The attributable burden of cancer, expressed as a count and percentage, was estimated for 2015 (the most recent year for which national-level cancer incidence data were available). Where possible, estimates were examined by sex, age group and province.

PARs are estimated using i) relative risks or odds ratios and ii) prevalence of an exposure. The resulting PAR is then multiplied by age, sex or province-specific incidence to obtain the associated count. More specifically:

1. **Relative risks and odds ratios:** relative risks (RRs) and odds ratios (OR) provide a measure of the extent to which an exposure is associated with increased (or decreased) risk of cancer. Whenever possible, RRs or ORs were obtained from the Canadian research literature. If no Canadian estimates were available, they were obtained from published meta-analyses or a meta-analysis was performed by the ComPARe study group using representative international studies.

2. **Exposure prevalence:** information on the percentage of the population exposed to a risk factor was obtained from national-level or region-specific data sources or both. For example, data sources included the Canadian Community Health Survey and the Canadian Health Measures Survey. For several environmental exposures, environmental monitoring data from sites in various parts of Canada were used. Where possible, exposure prevalence data were collected by sex, age and province.
    - **Latency period:** this term refers to the time between when an individual is exposed to a risk factor and when a cancer develops. Because of the latency period, historical exposure prevalence data were used. Defining the latency period was necessary for identifying the most appropriate year of prevalence data. When high-quality cohort studies on the exposure-cancer association were available, the latency period was defined by the average or median follow-up time between the exposure and cancer, and prevalence estimates dated around the midpoint of the range of potential latency periods were used. When cohort data were not available, a 10-year latency period was assumed and the closest available prevalence estimate was used. For

most infectious agents, latency was assumed to range from zero to 16 years, based on available data from studies of prevalent and incident cancer cases.

3. **Cancer incidence**: data on new cancer diagnoses (i.e. cancer incidence) by province, sex and age group were obtained from the Canadian Cancer Registry (CCR). CCR data included information on all cancers diagnosed in Canada between 1992 and 2015.* Only data on individuals aged 18 years and older were included in ComPARe, and the lower age cutoffs differed by type of exposure.

How the future attributable burden of cancer was estimated

The future, potentially avoidable, burden of cancer was estimated for 2018 to 2042. Where possible, estimates (percentage and counts) were obtained by sex and geographic region. More specifically:

1. First, trends in past exposure prevalence (typically determined based on longitudinal surveys) were extrapolated by year up to 2032. For exposures where there was limited prevalence data on historical trends, future prevalence was assumed to remain unchanged. The future prevalence estimates, together with RRs or ORs, were used to estimate future PARs (percentage).
2. Second, past trends in cancer incidence (from 1992 to 2015) were used to estimate cancer incidence to 2042. These projected counts were multiplied with the calculated percentage PARs to obtain future counts of attributable cases.

How future reductions in cancer due to interventions were estimated

To assess how the future burden of cancer could be reduced if the population's exposure to risk factors were modified, a range of prevention targets (also called counterfactuals) were considered.

To identify prevention targets, a literature search of population-based guidelines and recommendations was conducted, as well as a search of population-based interventions that have been shown to be effective in experimental studies. For example, prevention targets for smoking were estimated based on the World Health Organization's 2020 goal of a 30% reduction in the prevalence of tobacco use. Targets were then validated with public health experts. A full list of primary prevention targets can be found in Table 3.

In addition to literature-informed prevalence targets, targets based on fixed amounts of 10%, 25%, 50% and 100% reductions in prevalence were used for all exposure-cancer pairs. These targets were used to provide a range of scenarios that could be compared with respect to their impact on reducing cancer burden.

* The methodological framework described in a BMJ Open article indicated that incidence data in ComPARe were available to 2012, but more recent data (to 2015) were added after that article was published. The exception is Quebec: data were available to 2010 but were estimated for subsequent years up to 2015 for this province.

Tables

Table 1. Definitions used in ComPARe for risk factors

Risk factor		Exposure definition
Lifestyle	Alcohol	Having any number of drinks per day (13.5 g of ethanol per drink)
	Birth control pills*	Birth control pills taken ever, <1 year, 1–5 years, 5–10 years, ≥10 years
	Excess weight (overweight and obese)	Body mass index (BMI) ≥25 kg/m^2
	Hormone replacement therapy (HRT)*	HRT use ever, current use, use for <5 years and ≥5 years
	Low calcium	Less than 1000 mg per day (calcium intake from food in 24 hours)
	Low fruit	Less than 4 servings a day
	Low vegetable	Less than 4 servings a day
	Low vitamin D	Serum concentration of 25-hydroxy vitamin D is ≤50 nmol/L
	Physical inactivity	Moderately inactive: daily energy expenditure based on leisure time physical activity is ≥1.5 and <3.0 kcal/kg/day Inactive: daily energy expenditure based on leisure time physical activity is <1.5 kcal/kg/day
	Red and processed meat	Any consumption of red or processed meat
	Second-hand smoking (also called passive smoking)	Regularly exposed to tobacco smoke in their home, a vehicle or a public place
	Sedentary behaviour	Sedentary ≥6 hours per day during leisure time
	Tobacco smoking (also called active smoking)	Current smoker (smoked cigarettes daily or occasionally at the time of the interview) or former smoker (did not smoke at the time of the interview and had smoked more than 100 cigarettes in lifetime)
Environmental	Outdoor air pollution ($PM_{2.5}$)	Any exposure (risk per 10 µg/m^3)
	Disinfection by-products	>25–50µg/L and >50µg/L exposure to trihalomethanes in drinking water
	Indoor tanning	Ever use (lifetime)
	Residential radon	Any exposure (risk per 100 Bq/m^3)
	Sunburn	Ever (adult)
	Sunbathing	Ever (adult)

Risk factor		Exposure definition
Infectious agents	Epstein-Barr virus (EBV)	Having the virus
	Helicobacter pylori bacteria (*H. pylori*)	Having the virus
	Hepatitis B virus (HBV)	Having the virus
	Hepatitis C virus (HCV)	Having the virus
	Human herpesvirus type 8 (HHV8)	Having the virus
	Human papillomavirus (HPV)	Having the virus
	Human T-cell lymphotropic virus type 1 (HTLV)	Having the virus
Occupational†	Solar radiation	Low, moderate or high exposure as defined and estimated by CAREX Canada
	Silica (crystalline)	Low, moderate or high exposure as defined and estimated by CAREX Canada
	Asbestos	Ever exposed
	Diesel engine exhaust	Low, moderate or high exposure as defined and estimated by CAREX Canada
	Shiftwork	Having a work schedule of rotating shifts (including nights) or of permanent night shifts, as defined by CAREX Canada estimated based on the Survey of Labour and Income Dynamics (SLID) 1996
Grouped risk factors		
All risk factors		All exposures (lifestyle, environment, infections) listed above, not including occupational
Other infections		• Epstein-Barr virus (EBV) • Hepatitis B virus (HBV) • Hepatitis C virus (HCV) • *Helicobacter pylori* bacteria (*H. pylori*) • Human herpesvirus type 8 (HHV8) • Human T-cell lymphotropic virus 1 (HTLV)
Unhealthy eating		• Less than 1000 mg calcium per day • Less than 4 servings of fruit per day • Less than 4 servings of vegetables per day • Any red or processed meat

Risk factor	Exposure definition
Unhealthy living	• Excess weight • Physical inactivity • Sedentary behaviour
Joint risk factors	
Tobacco smoking and alcohol	Current smoker or former smoker and having any number of drinks per day (13.5 g ethanol per drink)
Excess weight and physical inactivity	BMI ≥ 25 kg/m^2 and daily leisure-time physical activity less than 1.5 kcal/kg/day
Sunburn, sunbathing, indoor tanning	Ever exposed for each exposure
Occupational solar radiation, creosotes, polycyclic aromatic hydrocarbons, mineral oils	Exposure was defined separately for each individual risk factor
Acid mist, asbestos, arsenic, beryllium, cadmium, chromium, diesel engine exhaust, second-hand smoke, ionizing radiation, nickel, polycyclic aromatic hydrocarbons, radon, silica, welding fumes, rubber production, work as a painter, art glass manufacturing	Exposure was defined separately for each individual risk factor
Aromatic amines, diesel engine exhaust, polycyclic aromatic hydrocarbons, tetrachloroethylene, work as a hairdresser or barber, work as a painter, rubber production	Exposure was defined separately for each individual risk factor

* Examined in ComPARe but not published

† This category of risk factors are examined in the Occupational Burden of Cancer study

Table 2. Definitions for cancer sites examined in ComPARe

Cancer	ICD-O-3* Topography (anatomical site)	ICD-O-3 Histology (appearance of cells under microscope)
Anal cancer	C21.0–C21.2, C21.8	8000–9049, 9056–9139, 9141–9589
Bile duct	C22.1, C24.0–C24.9, C26.8–C26.9, C48.0, C48.1–C48.2, C48.8	8000–9049, 9056–9139, 9141–9589
Bladder cancer	C67.0–C67.9	8000–9049, 9056–9139, 9141–9589
Breast cancer	C50.0–C50.9	8000–9049, 9056–9139, 9141–9589
Cervical cancer	C53.0–C53.9	8000–9049, 9056–9139, 9141–9589
Colorectal cancer	C18.0–C18.9, C26.0, C19.9, C20.9	8000–9049, 9056–9139, 9141–9589
Esophageal cancer	C15.0–C15.9	8000–9049, 9056–9139, 9141–9589
Eye cancer	C69.0–C69.9	8000–9049, 9056–9139 ,9141–9589
Gallbladder cancer	C23.9	8000–9049, 9056–9139, 9141–9589
Head and neck cancers†	C00.0–C00.9, C01.9–C02.9, C07.9–C08.9, C03.0–C03.9, C05.0–C05.9, C06.0–C06.9, C11.0–C11.9, C10.0–C10.9, C12.9, C13.0–C13.9, C09.0–C09.9, C14.0, C14.2–C14.8	8000–9049, 9056–9139, 9141–9589
Hodgkin lymphoma	C00.0–C80.9	9650–9667
Kaposi sarcoma	C00.0–C80.9	9140
Kidney cancer	C64.9, C65.9	8000–9049, 9056–9139, 9141–9589
Adult T–cell leukemia/lymphoma	C42.0, C77.0	9827
Acute myeloid leukemia	C00.0–C80.9	9840, 9861, 9865, 9866, 9867, 9869, 9871–9874, 9895–9897, 9898, 9910, 9911, 9920
Liver cancer	C23.9	8000–9049, 9056–9139, 9141–9589
Lung cancer	C34.0–C34.9	8000–9049, 9056–9139, 9141–9589
Melanoma	C44.0–C44.9	8720–8790
Mesothelioma	C00.0–C80.9	9050–9055
Multiple myeloma	C00.0–C80.9	9731–9732, 9734
Non–Hodgkin lymphoma	C00.0–C80.9 (all topographies excluding C42.0, C42.1, C42.4)	9590–9597, 9670–9729, 9735–9738, 9811–9818, 9823, 9827, 9837
Non–melanoma skin cancer	C44.0–C44.9	Basal cell carcinoma: 8901, 809–8094, 8097, 8098 Squamous cell carcinoma: 8070, 8074, 8075, 8078
Ovarian cancer	C56.9	8000–9049, 9056–9139, 9141–9589
Pancreatic cancer	C25.0–C25.9	8000–9049, 9056–9139, 9141–9589

Cancer	ICD-O-3* Topography (anatomical site)	ICD-O-3 Histology (appearance of cells under microscope)
Penile cancer	C60.0–C60.9	8000–9049, 9056–9139, 9141–9589
Prostate cancer	C61.9	8000–9049, 9056–9139, 9141–9589
Small intestine cancer	C17.0–C17.9	8000–9049, 9056–9139,9141–9589
Stomach cancer	C16.0–C16.9	8000–9049, 9056–9139,9141–9589
Thyroid cancer	C73.9	8000–9049, 9056–9139,9141–9589
Ureter cancer	C66.9	8000–9049, 9056–9139,9141–9589
Uterine cancer	C55.9	8000–9049, 9056–9139,9141–9589
Vaginal cancer	C52.9	8000-9049, 9056-9139,9141-9589
Vulvar cancer	C51.0	8000-9049, 9056-9139,9141-9589

*ICD-O-3=International Classification of Disease for Oncology, 3rd edition

† Oral cavity and pharynx (lip, tongue, salivary gland, floor of mouth, gum and other mouth, nasopharynx, oropharynx, hypopharynx, other [including tonsil])

Table 3. Prevention targets selected for estimating future burden of cancer

Risk factor		Published guideline or recommendation	Published target level
Lifestyle	Alcohol	Canada's Low-Risk Alcohol Drinking Guidelines: ≤ 15 drinks/week, ≤ 3 drinks/day for men; ≤ 10 drinks/week, ≤ 2 drinks/day for women. WCRF recommendation for cancer prevention: ≤ 2 drinks/day for men; ≤ 1 drink/day for women.	WHO Global NCD Action Plan 2013–2020: at least 10% relative reduction in the harmful use of alcohol as appropriate, within the national context by 2025 compared to 2010.
	Excess weight	Overweight: BMI 25 to < 30 kg/m^2 Obesity: BMI ≥ 30 kg/m^2	WHO Global NCD Action Plan 2013–2020: Halt the rise in obesity by 2025
	Low calcium	Health Canada: recommended dietary allowance per day – 1000 mg for adults 19–50 years, 1000 mg for men 51–70 years, 1200 mg for men 71+, and 1200 mg for women 51+. Suggested range: ≥ 1200 mg and < 2000 mg.	None
	Low fruit and vegetable consumption	≥ 5 servings/day WHO recommendations (category 2 intervention): As part of a healthy diet low in fat, sugars and sodium, WHO suggests consuming more than 400 g (5 portions) of fruits and vegetables per day to improve overall health and reduce the risk of certain non-communicable diseases (linked to target for control blood pressure and obesity). Health Canada recommendations: Men: 8–10 servings (age 19–50), 7 servings (age 51+); Women: 7–8 servings (age 19–50), 7 servings (51+)	None
	Low vitamin D	≥ 600 IU/day	Health Canada: recommended dietary allowance per day – 600 IU (15 mcg) for adults 19–70, 800 IU (20 mcg) for adults 71+. Suggested range: ≥ 600 IU and < 4000 IU.

	Physical inactivity	> 2.9 kcal/kg/day Canadian and WHO guideline: 150 minutes of moderate-to-vigorous activity/week.	WHO Global NCD Action Plan 2013-2020: a 10% relative reduction in prevalence of insufficient physical activity by 2025 compared to 2010.
	Red and processed meat	WCRF recommendation: Limit red meat and avoid processed meat. Public health goal: population average consumption of red meat to be no more than 300 g a week (≤ 3 servings/week), very little (if any) to be processed.	None
	Second-hand smoking	None	None
	Sedentary behaviour	National Institutes of Health's Physical Activity Guidelines for Americans (2018): Adults should move more and sit less throughout the day. Some physical activity is better than none. Adults who sit less and do any amount of moderate-to-vigorous physical activity gain some health benefits.	< 3 hours of sitting time in leisure time.
	Tobacco smoking	None	WHO Global NCD Action Plan 2013–2020: a 30% relative reduction in prevalence of current tobacco use in persons aged 15+ years by 2025 compared to 2010. A 3.7% reduction in prevalence based on research evidence.
Environmental	Outdoor air pollution ($PM_{2.5}$)	Canadian Environmental Quality Guidelines is 10ug/m^3 for annual mean $PM_{2.5}$ (Canadian Council of Ministers of the Environment 2009). WHO Air Quality Guideline for annual mean PM2.5 is 10ug/m^3 (WHO, 2005).	None
	Disinfection by-products	Guideline for Canadian Drinking Water Quality for total trihalomethanes is 100 ug/L.	None
	Indoor tanning	None	Health Canada recommends that people do not use indoor tanning devices.

			WHO recommends that people do not use indoor tanning devices.
	Residential radon	Health Canada residential radon guideline is 200Bq/m^3. WHO residential radon guideline is 100Bq/m^3 (WHO, 2009).	None
	Sunburn and sunbathing	None	Health Canada recommends using sun protection methods (limit time in the sun, watch UV index, wear protective clothing, use sunscreen, drink plenty of cool liquids) WHO recommends using sun protection methods (limit midday sun, watch UV index, use shade, wear protective clothing, use sunscreen)
Infectious agents	*Helicobacter pylori* bacteria (*H. pylori*)	5% and 35% reduction, where 5% is realistic and 35% the hypothetical ideal reduction.	Assumed 10%, 25%, 50% reductions
	Hepatitis B virus (HBV)	Set to represent a 10% and 20% increase in current vaccination coverage: Among adults: 50% and 60% Among those < 19 years: 80% and 90%	Assumed 10%, 25%, 50% reductions
	Hepatitis C virus (HCV)	40% reduction	Assumed 10%, 25%, 50% reductions
	Human papillomavirus (HPV)	Among vaccinated cohort: 80% and 90% vaccine coverage across Canada. Among women older than vaccinated cohort (aged 23+ but less than age 46) 15% and 25% coverage.	Based on HPV vaccine coverage for girls (range: 40–100%) and boys (range: 36–100%)

Glossary

Note that all terms have been defined in the context of this study.

attributable cases (also called excess attributable cases)	The number of cancer cases caused by a risk factor; the number of cancer cases that would have been avoided if exposure to the risk factor did not exist. Obtained by converting the PAR (%) into a count.
exposure (also called risk factor)	Any agent (e.g. alcohol, air pollution, HPV) that an individual may come into contact with. Sometimes a certain level of exposure is needed for the agent to be considered a risk factor (e.g. eating less than 4 servings of fruit per day is a risk factor for colorectal cancer).
exposure-cancer pair	Used to indicate causal associations between a risk factor and a specific cancer.
exposure prevalence	Percentage of a population exposed to a given risk factor.
incidence	Newly diagnosed cases of cancer in a given year.
joint effects	Used to describe instances when the effects of 2 or more risk factors are different from the effects of the individual risk factors.
latency period	Time period between when an individual is exposed to a risk factor and when the cancer is diagnosed.
modifiable risk factor	A risk factor that can be manipulated or changed to reduce its impact on developing disease. For example, age is not a modifiable risk factor, but tobacco smoking is a modifiable risk factor because the behaviour can be changed.
population attributable risk (PAR) (also called percent attributable)	The percentage of cancer cases in the population caused by exposure to a given risk factor.
potential impact fraction (PIF)	The future burden of cancer that could be prevented if exposure to a given risk factor was reduced.
prevention target (also called counterfactual)	Desired level of exposure prevalence that we aim to achieve through an intervention.
projected	Estimate of future rate of exposure prevalence or cancer incidence
relative risk (RR) and odds ratio (OR)	Measure of the extent to which an exposure is associated with an increased (or decreased) risk for cancer.
risk factor	A risk factor is anything that increases the chance of developing cancer.