CANCERS ATTRIBUTABLE TO INFECTIONS IN NORTH AMERICA

Karena Diana Volesky

A thesis submitted to McGill University in partial fulfillment of the requirements

for the degree of Doctor of Philosophy in Epidemiology

Department of Epidemiology, Biostatistics, and Occupational Health

McGill University, Montreal, Canada

April 2022

Contents

Abstract

Infections are a key contributor to cancer incidence, and among the most modifiable causes of cancer. The untapped potential for the prevention of carcinogenic infections makes quantifying their impact on cancer incidence a priority. Yet, the existing literature lacks comprehensive estimates on the impact that infections have on cancer incidence in North America.

The goal of this research was to assess the impact of seven infections on North American cancer incidence. Population attributable fractions were used to estimate the proportion of cancer incidence associated with a given infection. Calculating a population attributable fraction requires the prevalence of the infection in the general population and its relative risk associated with the cancer, or when mechanistic evidence permits, the prevalence of the infection within the cancer tissue. The prevalence of the infection in the general population for hepatitis B and C viruses (HBV, HCV) and *Helicobacter pylori* (*H. pylori*) were derived from North American population-based serosurveys. Meta-analytic methods were used to calculate the pooled measures of association (for HBV, HCV and *H. pylori*) and the prevalence in cancer cases (for Epstein-Barr virus [EBV], human papillomavirus [HPV], human herpesvirus type 8, and human T-cell lymphotropic virus type 1).

After analyzing 61 studies covering 20 different cancers, we found that 3.7% of the 189,530 cancers diagnosed among individuals aged ≥18 years in Canada in 2015 were attributable to infections. Analyzing 125 studies covering 26 different cancers, we found that of the 1,662,102 cancers diagnosed among individuals aged ≥20 years old in the United States in 2017, 4.3% were attributable to infections. HPV was the most important infectious cause of cancer accounting for 54.0% of the infection-attributable cancers in Canada and 53.8% in the United States.

Next, via a systematic review and meta-analysis, we estimated the fraction of incident cancers among individuals aged <20 years old in Europe and North America in 2020 that are attributable to EBV. We found that 2.6% of the estimated 42,654 incident cancers were attributable to EBV, of which 76.3% of cancers attributable to EBV were Hodgkin lymphomas.

Finally, we estimated the future burden of cancers caused by four major infections (HBV, HCV, *H. pylori* and HPV). The future burden was calculated by modelling: 10%, 25%, and 50% relative reductions in the prevalence of HBV, HCV and *H. pylori*, and different school-based HPV vaccination coverage levels (lower, current, higher) on Canadian cancer incidence by the year 2042. We found that

almost 16,000 cancers could be prevented in Canada from 2018 to 2042 with a 50% relative reduction in HBV, HCV and *H. pylori* prevalence and 80% HPV vaccine coverage of girls and boys.

While confirming the important impact infections have on cancer incidence in North America, these findings indicate that infections represent a key target for the development of prevention efforts (EBV) and the continuation or acceleration of current approaches to reduce the prevalence and associated cancer burden of HBV, HCV, *H. pylori*, and HPV.

Résumé

Les infections sont un facteur clé de l'incidence du cancer et comptent parmi les causes de cancer les plus faciles à modifier. Le potentiel inexploité de la prévention des infections cancérigènes rend prioritaire la quantification de leur impact sur l'incidence du cancer. Pourtant, la littérature existante manque d'estimations complètes de l'impact des infections sur l'incidence du cancer en Amérique du Nord.

L'objectif de cette recherche était d'évaluer l'impact de sept infections sur l'incidence du cancer en Amérique du Nord. Les fractions étiologiques du risque ont été utilisées pour estimer la proportion de l'incidence du cancer associée à une infection donnée. Pour calculer une fraction étiologique du risque, il faut connaître la prévalence de l'infection dans la population générale et son risque relatif associé au cancer ou, lorsque les preuves mécanistiques le permettent, la prévalence de l'infection dans le tissu cancéreux. La prévalence de l'infection dans la population générale pour les virus de l'hépatite B et C (VHB, VHC) et *Helicobacter pylori* (*H. pylori*) a été dérivée des études sérologiques de la population nord-américaine. Des méthodes de méta-analyse ont été utilisées pour calculer les mesures regroupées de l'association (pour le VHB, le VHC et *H. pylori*) et la prévalence dans les cas de cancer (pour le virus d'Epstein-Barr [VEB], le virus du papillome humain [VPH], le virus herpétique humain de type 8 et le virus humain T-lymphotrope de type 1).

Après avoir analysé 61 études couvrant 20 cancers différents, nous avons constaté que 3,7% des 189 530 cancers diagnostiqués chez les personnes âgées de ≥18 ans au Canada en 2015 étaient attribuables à des infections. Après avoir analysé 125 études couvrant 26 cancers différents, nous avons constaté que sur les 1 662 102 cancers diagnostiqués chez les personnes âgées de ≥20 ans aux États-Unis en 2017, 4,3% étaient attribuables à des infections. Le VPH était la plus importante cause infectieuse de cancer, représentant 54,0 % des cancers attribuables aux infections au Canada et 53,8% aux États-Unis.

Ensuite, par le biais d'une revue systématique et d'une méta-analyse, nous avons estimé la part des cancers incidents chez les individus âgés de moins de 20 ans en Europe et en Amérique du Nord en 2020 qui sont attribuables à l'EBV. Nous avons constaté que 2,6% des 42 654 cancers incidents estimés étaient attribuables à l'EBV, dont 76,3% de lymphomes Hodgkiniens.

Enfin, nous avons estimé la part future des cancers causés par quatre infections majeures (VHB, VHC, *H. pylori* et VPH). Le fardeau futur a été calculé par modélisation : Des réductions relatives de

10%, 25% et 50% de la prévalence du VHB, du VHC et de *H. pylori*, ainsi que différents niveaux de couverture vaccinale contre le VPH dans les écoles (inférieur, actuel, supérieur) sur l'incidence des cancers au Canada d'ici 2042. Nous avons constaté que près de 16 000 cancers pourraient être évités au Canada de 2018 à 2042 avec une réduction relative de 50% de la prévalence du VHB, du VHC et du H. pylori et une couverture vaccinale contre le VPH de 80% chez les filles et les garçons.

Tout en confirmant l'impact important des infections sur l'incidence du cancer en Amérique du Nord, ces résultats indiquent que les infections représentent une cible clé pour le développement d'efforts de prévention et la poursuite ou l'accélération des approches actuelles visant à réduire la prévalence et le fardeau du cancer associé au VHB, VHC, *H. pylori* et VPH.

Acknowledgements

To my thesis supervisor, Dr. Eduardo Franco, thank you for many years mentorship and for sharing your scientific wisdom. I appreciate the many opportunities you gave me to develop myself as an epidemiologist. Your drive and passion for your work is inspiring. It has been a privilege to conduct research on your team.

I was fortunate to have spent the first three years of my doctoral studies on a superstar team led by Dr. Darren Brenner and Dr. Christine Friedenreich – thank you both for the opportunities you provided me as a member of the ComPARe Study Team. I'm indebted to ComPARe Study Team members, Dr. Yibing Ruan and Ms. Abbey Poirier for the data analysis support they provided.

To my advisory committee members Dr. Stephen Walter, Dr. Allan Hildesheim, and Dr. Thomas O'Brien, I'm grateful for the insightful advice they provided on various elements of this work. Thank you to Ms. Sheila Bouten and Ms. Olga Tsyruk who contributed to the laborious exercise of extracting and verifying data. I'm also grateful for the contributions Dr. Samantha Morais and Dr. Eric Engels made to the second manuscript.

The Division of Cancer Epidemiology was a great place to work, and I'll always be thankful for the friendships I made there. A special thanks to Michel, Mariam, Cass, Talía, Farzin, Claudie, and Aaron for providing sound epidemiologic, career, and personal advice.

I was fortunate to have received a doctoral award jointly funded by the Fonds de recherche du Québec – Santé and the Fondation des étoiles; and for funding from the Canadian Cancer Society Partner Prevention Research Grant (to Dr. Franco) for the ComPARe Study.

Thank you to researchers, past and present, who conduct rigorous studies evaluating the relationships between infections and cancer. This work was only possible because those studies exist.

Florent, merci de toujours soutenir mon travail. J'ai la chance d'avoir à mes côtés un homme aussi attentif, compétent et calme.

To my family and dear friends – you are a blessing. I'm full of gratitude for your love and support. Mom, you made the enormous sacrifice of prioritizing us over your own PhD thesis, and we're grateful – this achievement is dedicated to you.

Contribution to Original Knowledge

The four manuscripts in this thesis represent original scholarship because manuscript #1 is the first study to estimate the individual and collective impact of seven different carcinogenic infections on cancer incidence in Canada and manuscript #2 is the first study to do so for the United States. Manuscript #3 is the first study to estimate the proportion and number of cancers among children that can be attributed to Epstein-Barr virus infection. Manuscript #4 is the first study to estimate how reductions in the prevalence of hepatitis B virus (HBV), hepatitis C virus (HCV), and *Helicobacter pylori* (*H. pylori*) could impact cancer incidence in Canada in the future.

Select specific original contributions are listed.

Manuscript #2 contains several methodological refinements and substantive additions not found in previous studies. For example, we used multiple imputation on HBV, HCV, and *H. pylori* prevalence estimates originating from National Health and Nutrition Examination Survey (NHANES) data. While several studies have addressed the underestimation of HCV prevalence estimates from NHANES data (HCV prevalence is underestimated due to the exclusion of groups with higher HCV prevalence from the NHANES sampling frame), none used multiple imputation to do so or provided more granular sex and age-group estimates. Further, we are the first to quantitatively address the underestimation of HBV and potential underestimation of *H. pylori* prevalence in NHANES data. Existing country-level studies estimating the impact of infections on cancer incidence did not include the role of EBV in diffuse large B-cell lymphoma (DLBCL), gastric carcinoma, or cancers diagnosed in children. Finally, to provide a more comprehensive estimate of *H. pylori*'s impact on cancer incidence, we included its protective effect in esophageal adenocarcinoma – a cancer not considered by previous studies.

Several studies have modelled the impact of specific interventions (e.g., HCV treatment, HPV vaccination coverage), in specific settings or populations, with cancer and its precursors, but manuscript #4 is the first to forecast the burden of all infection-attributable cancers by including the 12 cancers for which highly effective interventions exist (HBV, HCV, *H. pylori* and HPV).

Contribution of Authors by Manuscript

With guidance from Dr. Eduardo Franco and input from coauthors, I was the primary individual responsible for devising specific research questions, selecting infections (exposures) and cancers (outcomes), collecting data, performing analyses, interpreting the initial results, writing the first draft and subsequently revising each manuscript.

My committee members – Dr. Stephen Walter and Dr. Allan Hildesheim provided valuable feedback on manuscripts #2 and #3; Dr. Thomas O'Brien provided important direction and feedback on manuscript #2. ComPARe Study Group members, in particular – Dr. Yibing Ruan and Ms. Abbey Poirier, were consulted at various points in the development and conduct of manuscripts #1 and #4. I researched and wrote the literature review and discussion (Chapters 1 and 5, respectively) and Dr. Eduardo Franco reviewed and revised these chapters.

Manuscripts #1 and #4 originated from a larger project – the Canadian Population Attributable Risk of Cancer (ComPARe) Study. The ComPARe Study was a pan-Canadian project that 1. provided estimates of the proportion and number of cancer cases in Canada in 2015 due to 21 modifiable (or potentially modifiable) lifestyle, environmental and infectious agent risk factors and 2. quantified the annual number of incident cancer cases that would occur between 2018 and 2042 and the potential impact of prevention initiatives on that cancer incidence.

The ComPARe primary investigators were Dr. Darren Brenner (Co-PI, Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services), Dr. Christine Friedenreich (Co-PI, Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services), Dr. Stephen Walter (Department of Health Research Methods, Evidence, and Impact, McMaster University), Dr. Will King (Department of Public Health Sciences, Queen's University), Dr. Eduardo Franco (Division of Cancer Epidemiology, McGill University), Dr. Paul Demers (Occupational Cancer Research Centre), Dr. Paul Villeneuve (Department of Health Sciences, Carleton University), Dr. Prithwish De (Surveillance and Cancer Registry Department, Cancer Care Ontario), and Dr. Robert Nuttall (Health Quality Council of Ontario). Other ComPARe Study Team members were Dr. Yibing Ruan (Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services), Abbey Poirier (Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services), Xin Grevers (Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services), Dr.

Mariam El-Zein (Division of Cancer Epidemiology, McGill University), Dylan O'Sullivan (Department of Public Health Sciences, Queen's University), Tasha Narain (Department of Public Health Sciences, Queen's University), Priyanka Gogna (Department of Public Health Sciences, Queen's University), Dr. Leah Smith (Canadian Cancer Society), Elizabeth Holmes (Canadian Cancer Society), Zeinab El-Masri (Cancer Care Ontario).

Specific contributions to each manuscript are listed.

Manuscript #1: Volesky KD, El-Zein M, Franco EL, Brenner DR, Friedenreich CM, Ruan Y, ComPARe Study Team. Cancers attributable to infections in Canada.

KDV, ELF, DRB, and CMF conceived the study. KDV performed the literature search, screened records for eligibility, and extracted the data. KDV and YR developed the analytical design; KDV analyzed the data for four infections, and for three infections KDV analyzed the data with assistance from YR. KDV wrote the first draft of the manuscript, and all authors revised the manuscript.

Manuscript #2: Volesky KD, Morais S, Walter SD, O'Brien TR, Hildesheim A, Engels EA, El-Zein M, Franco EL. Cancers Attributable to Infections in the United States in 2017.

KDV and ELF conceived the study. With input from SM, TOB, AH, and EAE, KDV determined which cancers warranted inclusion. KDV and SDW developed the analytical design. KDV and SM collected, extracted, and analyzed data. MZ linked the population-based datasets. KDV wrote the first draft of the manuscript, and all authors reviewed the manuscript.

Manuscript #3: Volesky KD, Tsyruk O, Hildesheim A, Walter SD, El-Zein M, Friedenreich CM, Brenner DR, Franco EL. Epstein-Barr virus and cancer among children and adolescents in Europe and North America: a systematic review, meta-analysis, and attributable burden estimation.

KDV and ELF conceived the study. KDV, AH, SDW, and ELF developed the protocol which was revised by all authors. KDV performed the literature search and screened records for eligibility. OT and KDV extracted the data. KDV performed the analyses. KDV wrote the first draft of the manuscript, and all authors revised the manuscript.

Manuscript #4: Volesky KD, El-Zein M, Franco EL, Brenner DR, Friedenreich CM, Ruan Y, ComPARe Study Team. Estimates of the future burden of cancer attributable to infections in Canada.

KDV, ELF, DRB, and CMF conceived the study. KDV performed the literature search, screened records for eligibility, and extracted the data. KDV analyzed the data for HPV and for the remaining three infections, KDV analyzed the data with assistance from YR. KDV wrote the first draft of the manuscript, and all authors revised the manuscript.

Statement of Financial Support

The first year of my doctoral studies (2016–2017) was financed by a stipend originating from a grant for the Canadian Population Attributable Risk of Cancer (ComPARe) Study for which Dr. Eduardo Franco was a Co-PI on. The ComPARe Study was supported by a Canadian Cancer Society Partner Prevention Research Grant (#703106). The following three years, I was supported by a doctoral award jointly provided by Fonds de recherche du Québec – Santé (FRQS) and Fondation des étoiles (2017–2020), funding that was generously topped up by a stipend from Dr. Franco. The following year, I was funded by a stipend from Dr. Eduardo Franco's portfolio of research grants (2020–2021). I attended and presented research findings from different analyses from several of the manuscripts contained within this thesis at several domestic and international conferences, these presentations were funded by funds from the ComPARe Study and by travel awards.

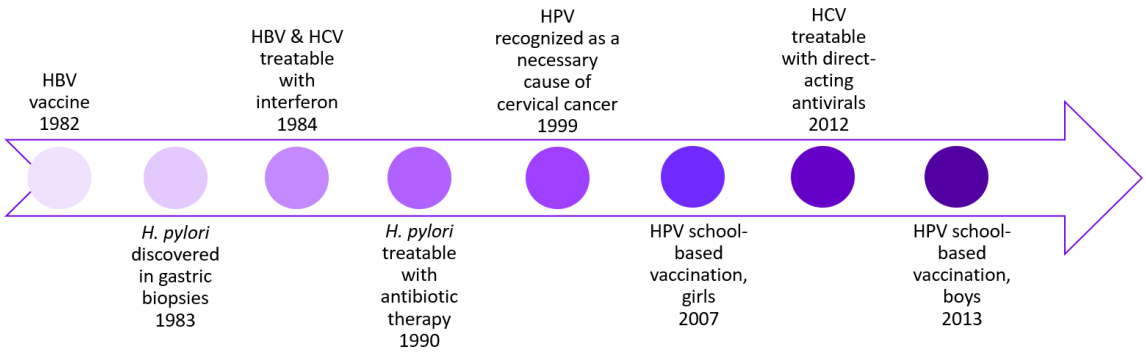
Abbreviations

AC	Attributable cases
BL	Burkitt lymphoma
CHMS	Canadian Health Measures Survey
CI	Confidence interval
ComPARe	The Canadian Population Attributable Risk of Cancer Study
DLBCL	Diffuse large B-cell lymphoma
EBER	EBV-encoded small RNAs
EBV	Epstein-Barr virus
EIA	Enzyme immunosorbent assay
ELISA	Enzyme-linked immunosorbent assay
ENKTL	Natural killer T-cell lymphoma
HBsAg	Hepatitis B surface antigen
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HHV-8	Human herpesvirus, type 8
HIV-1	Human immunodeficiency virus, type 1
HL	Hodgkin lymphoma
HNC	Head and neck cancer
H. pylori	*Helicobacter pylori*
HPV	Human papillomavirus
HTLV-1	Human T-cell lymphotropic virus, type 1
IARC	International Agency for Research on Cancer
IICC	International Incidence of Childhood Cancer
ISH	*In situ* hybridization
JBI	Joanna Briggs Institute
LMP	Latent membrane protein
MALT	Mucosa-associated lymphoid tissue
NHANES	National Health and Nutrition Examination Survey
NHL	Non-Hodgkin lymphoma
NLPHL	Nodular lymphocyte-predominant Hodgkin lymphoma
OR	Odds ratio
PAF	Population attributable fraction
PAR	Population attributable risk
PEL	Primary effusion lymphoma
PIF	Potential impact fraction
PLWH	People living with HIV
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RR	Relative risk
RSE	Relative standard error
SEER	Surveillance, Epidemiology, and End Results
US	United States

Chapter 1: Introduction

Eleven infections, including seven viruses, three parasites and one bacterium, are recognized as established carcinogens.[1] Together, these 11 infections are capable of leading to more than 20 different types of cancer.[1,2] The recognition of the role of infections in human cancer, started with the discovery of the first human tumour virus, Epstein-Barr virus (EBV) in the mid-1960s.[3] In the decades that followed, there were other important discoveries in this field, from the identification of the bacterium *Helicobacter pylori* (*H. pylori)* in gastric biopsies in 1983 to the recognition that human papillomavirus (HPV) is a necessary cause of cervical cancer in 1999.[4-6] Paralleling these discoveries were advances in preventing and treating carcinogenic infections (**Fig. 1**).[7-10]

Fig. 1. Examples of major milestones, dates are relevant to Canada



HBV = hepatitis B virus, HCV = hepatitis C virus, *H. pylori* = *Helicobacter pylori*, HPV = human papillomavirus

Despite established relationships between certain infections and cancer, and past and current efforts to prevent and treat carcinogenic infections, there remains unrealized potential for primary and secondary prevention of carcinogenic infections. The existing literature lacks comprehensive estimates on the impact that infections have on cancer incidence in Canada, the United States and among children. Yet, such estimates contribute to the evidence needed to prioritize the development and implementation of strategies aimed at reducing the prevalence of certain carcinogenic infections.

Research Objectives

The principal goal of my manuscript-based doctorate research is to assess the impact of infections on North American cancer incidence by comprehensively estimating the proportion and number of cancers caused by infections. The specific objectives in pursuit of this goal are to estimate

1. Among individuals aged 18 and older, the percentage and number of incident cancers attributable to infections in Canada in 2015.
2. Among individuals aged 20 and older, the percentage and number of incident cancers attributable to infections in the United States in 2017.
    a. Among individuals aged 19 and younger, the percentage and number of incident cancers attributable to EBV in 2017.
3. Among individuals aged 19 and younger, the percentage and number of incident cancers attributable to EBV in Europe and North America in 2020.
4. The future burden of infection-associated cancer by the year 2042 in Canada by modelling the impact of:
    a. Relative reductions in hepatitis B virus, hepatitis C virus, and *Helicobacter pylori* infection prevalence, and
    b. Lower, current, and higher levels of school-based human papillomavirus vaccination coverage.

Organization of Dissertation

This five-chapter manuscript-based dissertation starts with an introduction chapter (Chapter 1) and ends with a chapter discussing the main findings, limitations, and implications of this research (Chapter 5). Chapters 2 to 4 contain empirical work quantifying the burden of cancer due to infections: among adults residing in Canada and the United States (Chapter 2), among children and adolescents residing in North America or Europe (Chapter 3), and among adults in Canada in the future (Chapter 4). Chapter 5 contains the discussion. Supplementary material appears at the end of the manuscript it is connected to. Bridging text connects the different chapters of this thesis. All citations appear at the end of this dissertation.

Ethics

Ethics approval was granted for the Canadian portion of this project by the Health Research Ethics Board of Alberta – Cancer Committee (HREBA.CC-14-0220_REN4) and McGill University granted an exemption. Since the US and European portions utilize publicly available data, they are also exempt from ethics approval; however, consent disclosure forms have been processed for access to cancer incidence data (US only) and for some of the data requested directly from original study authors.

Literature Review

Cancer and its causes

In 2019, cancer was the leading cause of death in Canada, and in the United States (US) it was the second leading cause after heart disease.[11,12] In fact, cancer has been the leading cause of death in Canada for the last two decades.[12] Recent estimates indicate that the lifetime probability of developing cancer is 44% for males and 43% for females in Canada, and 41% for males and 39% for females in the US.[13,14] Since cancer represents a major public health issue in Canada and the US, quantifying the proportion of cancers that could potentially be adverted is an important component in addressing the cancer burden.

Causes of cancer can be broadly categorized into four factors: genetic, lifestyle, environmental, and infections. Since lifestyle, environmental and infectious agent risk factors are potentially modifiable, they have been the focus of country-level studies estimating the combined impact of these factors on cancer incidence.[15-19] Such estimates quantify the potential preventability of cancer by reporting the number of incident cancer cases that could in theory be avoided through changes in modifiable lifestyle, environmental, and infectious agent risk factors. For example, lifestyle, environmental and infectious agent risk factors were estimated to account for 32% of incident cancers in Australia in 2010,[15] 38% of cancers diagnosed in the United Kingdom (UK) in 2010,[16] 42% of cancers in the US in 2014,[17] 33% of cancers in Canada in 2015,[18] and 41% of cancers in France.[19] The top contributor to cancer incidence in these countries was tobacco smoking, followed by either inadequate diet (Australia), excess body weight (US and UK), physical inactivity (Canada), or alcohol consumption (France).[15-18] In each country, the combined impact of infections places it in the top 10 most important contributors to cancer incidence.[15-19]

Infections are among the most modifiable (or avoidable) cancer risk factors with highly effective interventions – vaccination for HBV and HPV confer >95% efficacy when administered prophylactically,[20,21] >90% of HCV infections are curable with highly active direct acting antivirals,[22,23] and *H. pylori* is treatable with antibiotic therapy.[7] The aforementioned interventions seem to have higher public acceptability and compliance, compared to other interventions aimed at reducing cancer risk such as encouraging exercise and weight loss.

Infections capable of causing cancer

The International Agency for Research on Cancer (IARC) classifies exposures, including infections, as carcinogenic (Group 1),[1] probably (Group 2A) or possibly carcinogenic (Group 2B), not classifiable (Group 3),[24] based on well defined criteria.[1,25,26] To date, IARC has classified 121 agents as carcinogenic (Group 1),[27] 11 of which are infections (**Table 1**).

Table 1. International Agency for Research on Cancer classification of infectious agents[28]

Agent	Type of organism	Group[a]
Hepatitis B virus	Virus	1
Hepatitis C virus	Virus	1
Helicobacter pylori	Bacterium	1
Epstein-Barr virus/human herpesvirus type 4	Virus	1
Human papillomavirus (HPV) genotypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59	Virus	1
Kaposi sarcoma herpesvirus/human herpesvirus type 8	Virus	1
Human T-cell lymphotropic virus type I	Virus	1
Human immunodeficiency virus type 1	Virus	1
Opisthorchis viverrini	Metazoan	1
Schistosoma haematobium	Metazoan	1
Clonorchis sinensis	Metazoan	1
HPV genotype 68	Virus	2A
Merkel cell polyomavirus	Virus	2A
Malaria (*Plasmodium falciparum*)	Protozoan	2B
BK polyomavirus	Virus	2B
JC polyomavirus	Virus	2B
HPV genotypes 26, 30, 34, 53, 66, 67, 69, 70, 73, 82, 85, 97	Virus	2B
HPV genotypes 5 and 8 (in patients with epidermodysplasia verruciformis)	Virus	2B
Human immunodeficiency virus type 2	Virus	2B
Schistosoma japonicum	Metazoan	3
HPV genus beta (except types 5 and 8) and genus gamma	Virus	3
HPV types 6 and 11	Virus	3
Human T-cell lymphotropic virus type II	Virus	3
Hepatitis D virus	Virus	3
SV40 (simian virus) polyomavirus	Virus	3
Schistosoma mansoni	Metazoan	3
Opisthorchis felineus	Metazoan	3

[a] 1: carcinogenic, 2A probably carcinogenic, 2B possibly carcinogenic, 3: not classifiable

The IARC last reassessed the carcinogenicity of Group 1 infections and associated cancers in February 2009 and published its findings in monograph volume 100B published in 2012.[1,2] The discovery of the Merkel cell polyomavirus genome in Merkel cell carcinoma (a rare type of skin cancer) biopsies in 2008,[29] prompted IARC to evaluate Merkel cell polyomavirus, among three other polyomaviruses (Simian virus 40, BK and JC virus) and malaria, in February 2012.[30] The findings of this evaluation were published in 2014.[25,30] IARC also classifies cancers/cancer sites as having 'sufficient'

[1] Group 1 is the category that includes agents assessed in scientific studies that have demonstrated "Sufficient evidence of carcinogenicity in humans." or "Evidence less than sufficient in humans but sufficient in experimental animals and strong evidence that in exposed humans the agent acts through a relevant carcinogenic mechanism."

or 'limited' evidence for the role of the exposure. More than 10 years have elapsed since IARC last updated its evaluation of the spectrum of infectious agents. There is increasing evidence that infections may cause more cancers than those listed as having sufficient evidence, such as gastric carcinoma (EBV) and diffuse large B-cell lymphoma (DLBCL, EBV).

How infections can cause cancer

Infections can lead to cancer through direct and/or indirect mechanisms.[31] EBV, HPV and HHV-8 act directly through genome integration or interference with genetic control of cellular proliferation.[31] As direct carcinogens, the virus is present in each cancer cell and expresses one or more transcripts to maintain the cancer cell phenotype.[31] For example, EBV infects and replicates in oral lymphoepithelioid tissue, then persists as a latent infection within B-cells.[32] Notably, several EBV related cancers are of B-cell origin – Hodgkin and Burkitt lymphoma, and DLBCL. The EBV viral genome is present and transcriptionally active in each tumour cell, implying a continuing role for the virus in tumor growth.[32] HPV's mechanism of oncogenesis is the partial integration of its viral genome into host DNA resulting in the overexpression of early viral proteins (E6 and E7) which can inactivate tumour suppressors, TP53 and Rb gene products, respectively. Overexpression of E6 and/or E7 can lead to uncontrolled cell proliferation and immortalization.

Infections can cause cancer indirectly through long-term inflammation leading to changes in immune cells producing inflammatory mediators that can cause cancer (HBV, HCV, *H. pylori*), or through down-regulation of the immune system (human immunodeficiency virus [HIV]). For example, HBV integrates into the genomes of cancer cells (direct action), and indirectly through inflammation of the liver leading to liver cirrhosis – the most important risk factor for hepatocellular carcinoma development.

Each agent's specific mechanisms of carcinogenesis can be found in **Table 2.**

Brief background on the infections

Here, several key features of the carcinogenic infections included in this thesis are highlighted. First, several of the cancer-associated infections are common in healthy populations, such as EBV, HPV, and *H. pylori*. For example >90% of adults globally are infected with EBV,[33,34] and ~50% of the world's population harbour *H. pylori*.[35] Second, there is typically a long latency period between initial infection and cancer development (e.g., ~25 years between initial HCV infection and hepatocellular carcinoma development and ~50 years between HTLV-1 and adult T-cell leukemia/lymphoma).[1]

Third, only a small proportion of individuals with an infection will develop cancers; for example, in Western countries about 1–2% of those with *H. pylori* infection are expected to develop gastric cancer.[36,37] Fourth, the immune system in 'healthy' populations often clears or keeps the infection under control; however, those with compromised immune surveillance (i.e., people living with human immunodeficiency virus [HIV] or solid organ transplant recipients) are less likely to clear or control the infection.[38] Fifth, the oncogenicity can vary according to genotype/strain (e.g., HPV genotype 16 and *H. pylori* CagA+ gene strain are more potent carcinogens compared to other genotypes/strains).[39-41] Sixth, infections included in this thesis have been shown to have strong relationships with their respective cancer(s); for example, studies report odds ratios (ORs) or relative risks (RRs) of >20 for each HBV and HCV and hepatocellular carcinoma,[42] *H. pylori* and non-cardia gastric carcinoma RR>10,[45,46] HPV16 and oropharyngeal cancer OR>10.[47-49] Finally, cofactors are involved in infection associated cancers. For example, *H. pylori* is the major cause of gastric cancer, yet genetic, lifestyle (smoking, body fatness, process meat), environmental (e.g., rubber manufacturing) and even another infection (EBV) are also associated with gastric cancer (**Fig. 2**).[43,44] Another example is HPV and cervical cancer. While HPV is a necessary cause of cervical cancer (100% of cervical cancers are attributable to HPV),[5] there are several cofactors in the carcinogenesis of cervical cancer such as immunosuppression (HIV and solid organ transplantation), tobacco smoking, and even another infection – chlamydia.[50-53]

Fig. 2. Gastric cancer cofactors[43,44]

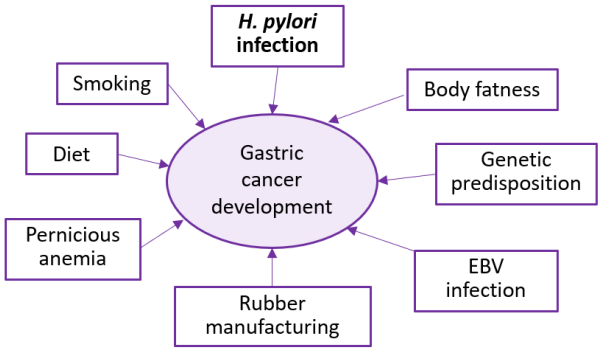


Table 2 contains a summary of select features of the infections included in this thesis. Please see the supplements of manuscripts #1 and #2 for additional background on the infections.

Table 2. Background on the carcinogenic infections included in this thesis research

Infection	Group (genome)	Carcinogenic mechanism(s)[b] *From Bouvard 2009*	Main transmission route(s)	Cancers with *sufficient* evidence[c]	Cancers with *limited* evidence[c]
Hepatitis B virus *chronic infection*	Hepadnavirus (3 Kb DNA)	Inflammation Liver cirrhosis Chronic hepatitis	Blood and other body fluids	Hepatocellular carcinoma	Cholangiocarcinoma Non-Hodgkin lymphoma
Hepatitis C virus *chronic infection*	Flavivirus (10 Kb RNA)	Inflammation Liver cirrhosis Liver fibrosis	Blood and less commonly through other body fluids	Hepatocellular carcinoma, non-Hodgkin lymphoma	Cholangiocarcinoma
Helicobacter pylori	*Helicobacter* (1.14 Mb DNA)	Inflammation Oxidative stress Altered cellular turn-over and gene expression Methylation Mutation	Oral/fecal	Non-cardia gastric carcinoma, low-grade B-cell MALT gastric lymphoma	None
Epstein-Barr virus[d]	Gamma-herpesvirus (~170 Kb DNA)	Cell proliferation Inhibition of apoptosis Genomic instability Cell migration	Oral/saliva	Burkitt lymphoma Hodgkin lymphoma, extranodal natural killer T-cell lymphoma - nasal type, nasopharyngeal carcinoma, immune suppression-related non-Hodgkin lymphoma	Gastric carcinoma Lymphoepithelioma-like carcinoma
Human papillomavirus	Papillomaviridae (8 Kb DNA)	Immortalisation, Genomic instability Inhibition of DNA damage response Anti-apoptotic activity	Mucosal/skin-to-skin	Cancers of the cervix, anus, penis, vagina, vulva, oropharynx, tonsil, and oral cavity	Laryngeal carcinoma
Human herpesvirus type 8[e]	Gamma-herpesvirus (~140 Kb DNA)	Cell proliferation, Inhibition of apoptosis Genomic instability Cell migration	Oral/saliva	Kaposi sarcoma, primary effusion lymphoma	Multicentric Castleman's disease
Human T-cell lymphotropic virus type 1	Retrovirus (10 Kb RNA)	Immortalisation and transformation of T cells	Blood and other body fluids, including breast milk	Adult T-cell leukemia/lymphoma	None
Human immunodeficiency virus type 1	Retrovirus (10 Kb RNA)	Immunosuppression (indirect action)	Blood and other body fluids	Kaposi sarcoma, non-Hodgkin lymphoma, Hodgkin lymphoma, cancer of the cervix, anus, conjunctiva	Cancer of the vulva, vagina, penis, nonmelanoma skin cancer, hepatocellular carcinoma

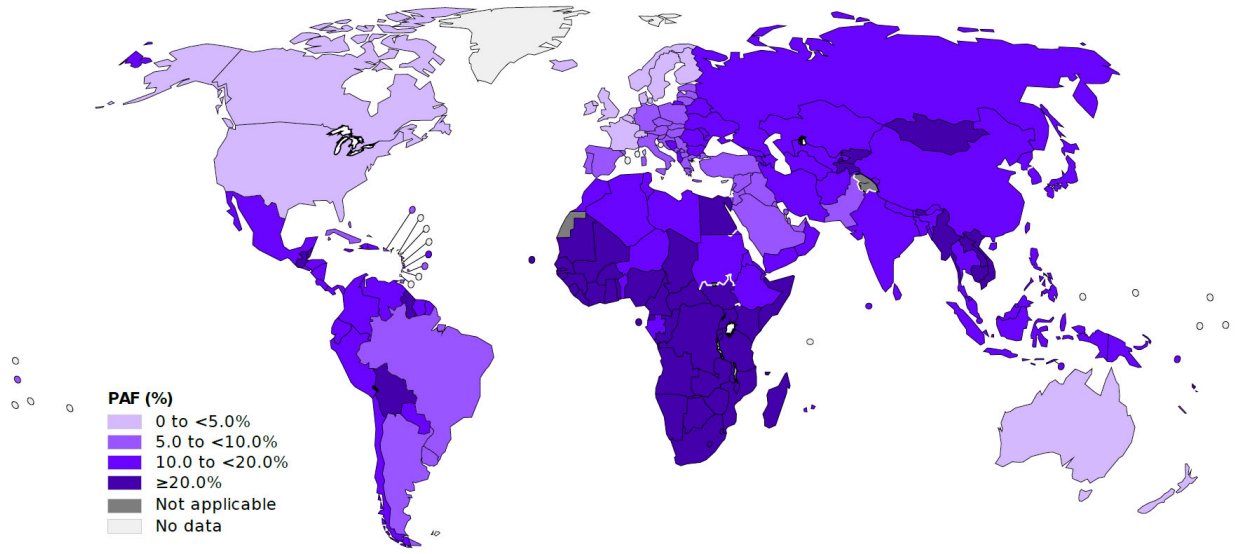
MALT = mucosa-associated lymphoid tissue

a. Included infections have been categorized by IARC as Group 1 carcinogens.
b. Carcinogenic mechanisms were taken from Bouvard 2009.[2]
c. Cancer sites were categorized by IARC as having *sufficient* or *limited* evidence.
d. Epstein-Barr virus is also referred to as human herpes virus, type 4.
e. Human herpesvirus, type 8 is also referred to as Kaposi sarcoma virus.

Impact of infections on cancer incidence

Globally, 13% of all cancers (excluding nonmelanoma skin cancers) were attributable to infections in 2018 with large regional variations observed (**Fig. 3**).[54-56] The proportion of infection-attributable cancers in 2018 was lowest in Norway and Sweden (3%) and highest in Mongolia and Mozambique (49%).[57]

Fig. 3. Percentage of all cancer cases among both sexes in 2018 attributable to infections, by country[57]



In Canada and the US, 3.6% and 4.8% of cancers were due to infections, respectively.[57] The global analysis combined infection prevalence for regions comprising several countries (i.e., regions with a low, medium and high prevalence of infections). Given infection prevalence varies by region, country-specific data including data obtained from population-based studies are important for calculating improved estimates of the impact of infections on cancer incidence. Since vaccines and treatments are infection-specific, estimating the proportion of cancers attributable to each infection provides an essential assessment of the cancer burden due to infections. While existing global analyses are of immense value[54,56,57] it is not feasible for these studies to obtain the most relevant data for each country/infection/sex/age group. For this reason, the limitations of performing a global analysis can be mitigated by conducting the country-level analyses shared here.

Since 2000, population attributable fraction (PAF) analyses of infections as a cause of cancer (year cancer incidence was applied to), have been published for several countries, including Australia

(2010),[58] Brazil (2020),[59] China (2005 and 2013),[60,61], France (2000 and 2015),[62,63] Italy (2018),[64] Korea (2007),[65] the Netherlands (2003),[66] the United Kingdom (2010 and 2015),[16,67] and the US (2014).[17] However, several of these analyses did not attempt to find region specific PAF inputs, and instead extracted existing PAF estimates from global PAF analyses.[61,64,66] Some teams attempted to find regionally applicable estimates but lacked the data to do so (e.g., Nigeria).[68] Given how the prevalence of carcinogenic infections, such as the hepatitis viruses, *H. pylori*, and EBV prevalence in lymphoma cancer tissues can vary across regions, prevalence estimates from one country are not necessarily applicable to another country. Hence, the need to obtain the most regionally applicable data on infection prevalence in the population and cancer cases to reliability estimate the PAF.

Overview of methods

The central method used in this research is the calculation of population attributable risks or fractions (PARs, PAFs, in this thesis the two terms are used interchangeably). The PAF quantifies the proportion of disease that could be avoided if the exposure were eliminated from the population.[69] PAFs are easy to interpret because they can be expressed as the percent of cases attributable to the exposure. PAFs can also be compared across a variety of modifiable risk factors to assess the attributable burden and combined to give an overall PAF for a group of exposures.[70] The central assumption of a PAF is that a causal relationship exists between the exposure and outcome. Another assumption is that of a counterfactual where the exposure could be eliminated/mitigated. The PAF was originally developed by Levin in 1953.[69]

$$PAF = \frac{Pe(RR-1)}{1+Pe(RR-1)}$$

Pe is the prevalence of the exposure in the general population
RR is the relative risk

The Pe is often obtained from a population-based survey and the RR from a meta-analysis. In practice, the RR is often substituted for the OR (subject to data availability). This formula assumes no confounding in the measure of association between exposure and disease.[69] The confidence intervals for Levin's formula can be calculated with the formula:

$$var[\ln(1-PAR)] = PAR^2\left[\frac{1}{P^2}varP + (\frac{RR}{RR-1})^2 var[\ln RR] + \frac{2}{P}(\frac{RR}{RR-1})cov(P, lnRR)\right]$$

Where *var* and *cov* are the variance and covariance, respectively.[71]

Alternatively, Miettinen's formula may be preferred when using adjusted measures of association.[72]

$$PAF = Pc * \frac{RR - 1}{RR}$$

Pc is the prevalence of the exposure in cases
RR is the relative risk

Miettinen's formula uses the exposure prevalence among cases, which is available within individual study data but not available for the target population.[72] For this reason, Levin's formula is often utilized with measures of association adjusted for confounders but with the caveat that is it an approximation of the PAF.

Another widely used formula to attributable cancers to certain infections such as EBV or HPV,[56,73] is: Pc ≈ PAF. There are two situations in which it is appropriate to approximate the PAF by Pc. First, when the measure of association is very high (i.e., >20) such that the attributable fraction in the exposed approaches 1.0, at which point the prevalence in cases approximates the PAF. Second, when mechanistic evidence supports that the detection of the infection in cancer cases infers that the cancer is attributable to the infection. Often, the prevalence in cancer cases is the prevalence of the infection within the cancer tissue/cells and detection of the infection is by suitable, ideally gold standard methods. A limitation of this method is the potential to overestimate the PAF.

Ideally, the PAF would be estimated from lifetime follow-up of a cohort of exposed and non-exposed people in the population of interest. Though the lifetime cohort method is not feasible, several cohorts have generated PAFs for a specific population and time and often a single exposure or cancer.[74,75] However, such studies are not designed to address the overall impact of infections on a nation's cancer incidence. For this reason, alternate approaches (such as the formulae presented above) are required.

An extension of the PAF is the potential impact fraction (PIF). A PIF can quantify the proportion of disease that could be avoided under counterfactual scenarios.[76]

$$PIF = \frac{(\mathrm{P} - \mathrm{P}^*)(\mathrm{RR} - 1)}{\mathrm{P}(\mathrm{RR} - 1) + 1}$$

P is the pre-counterfactual infection prevalence
P* is the post-counterfactual infection prevalence
RR is the relative risk

The PIF is then used to estimate the number of prevented cases in a given year by

$$PC_i = I_i \times PIF$$

$\boldsymbol{I_i}$ is the projected cancer incidence in year ***i***

Since PAFs and PIFs can change over time due to changing exposure prevalence or improvements in the techniques to measure exposures, among other reasons, they require regular updating with the most pertinent exposure and magnitude of risk information. Although calculating PAF/PIF estimates can be straightforward, selecting appropriate data for their inputs requires a thorough understanding of epidemiological and substantive features of the exposure and outcomes under study. Ideally, data for the PAF/PIF calculations are obtained via transparent and reproducible methods. More data permits more detailed analyses (by sex, age group, cancer subtype, and geographical area), thus emphasizing the need for comprehensive methods to obtain all relevant data.

Chapter 2: Current Burden of Infection Attributable Cancers in North America

Manuscripts #1 (Canada) and #2 (US) form the core of this thesis as they address the overarching goal of this thesis: to assess the impact of infections on North American cancer incidence. Both manuscripts employ many of the methods foundational to PAF work on infections and cancer and thus are similar in terms of their content. While data were collected for the Canadian study up to mid-2018, data were collected for the US study up to the end of 2021. During that time, the link between EBV and DLBCL and EBV and gastric carcinoma gained greater acceptance, and thus these two additional associations are included in the US manuscript but not the Canadian one.

Manuscript #1: Cancers attributable to infections in Canada

This manuscript includes estimates of the impact that seven Group 1 infections had on cancer incidence in Canada in 2015. This assessment was part of the ComPARe Study and helped fulfill the first of two goals of the ComPARe study: to provide estimates of the proportion and number of cancer cases in Canada in 2015 due to 21 modifiable (or potentially modifiable) lifestyle, environmental and infectious agent risk factors. Together, the 21 risk factors were responsible for 33.0% of the 2015 cancer burden among adults.[18] The manuscript was published as part of a special issue on the burden of cancer in Canada, in the journal *Preventive Medicine* in the spring of 2019. The published version of this manuscript can be found in the appendix.

Cancers attributable to infections in Canada

Karena D. Volesky[1,2]
Mariam El-Zein[1]
Eduardo L. Franco[1,2]
Darren R. Brenner[3]
Christine M. Friedenreich[4]
Yibing Ruan[4]
On behalf of the ComPARe Study Team[5]

1. Division of Cancer Epidemiology, Gerald Bronfman Department of Oncology, McGill University, Montréal, Québec, Canada
2. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada
3. Departments of Oncology and Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
4. Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada
5. Additional ComPARe Study Team members: **Stephen Walter**, Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada; **Will King**, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; **Paul Demers**, Occupational Cancer Research Centre, Toronto, Ontario, Canada; **Paul Villeneuve**, Department of Health Sciences, Carleton University, Ottawa, Ontario, Canada; **Abbey Poirier**, Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada; **Prithwish De**, Cancer Care Ontario, Toronto, Ontario, Canada; **Leah Smith**, Canadian Cancer Society, Toronto, Ontario, Canada; Elizabeth Holmes, Cancer Care Ontario, Toronto, Ontario, Canada; **Dylan O'Sullivan**, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; **Zeinab El-Masri**, Cancer Care Ontario, Toronto, Ontario, Canada; **Robert Nuttall**, Canadian Cancer Society, Toronto, Ontario, Canada; **Sheila Bouten**, Division of Cancer Epidemiology, McGill University, Montréal, Québec, Canada; **Elizabeth Holmes**, Canadian Cancer Society, Toronto, Ontario, Canada; **Tasha Narain**, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; **Priyanka Gogna**, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada

This manuscript is presented in the format it is published in the journal, Preventive Medicine, with minor modifications to appear more consistent with the other manuscripts contained in this thesis.

Volesky KD, El-Zein M, Franco EL, Brenner DR, Friedenreich CM, Ruan Y. Cancers attributable to infections in Canada. Prev Med. 2019;122:109-17.

Acknowledgements: Karena D. Volesky is supported by a doctoral award from the Fonds de recherche du Québec – Santé (FRQS) and the Fondation des étoiles. Darren Brenner holds a Canadian Cancer Society Prevention Capacity Development Award (#703917). Christine Friedenreich was supported by an Alberta Innovates Health Senior Scholar Award and by the Alberta Cancer Foundation Weekend to End Women's Cancers Breast Cancer Chair. We thank Shemay Lee for assisting us with data management.

Competing interests: None

Disclosure: E.L.F. has served as occasional consultant to companies involved with HPV diagnostics and vaccination (Merck, GSK, Roche, and BD). His institution has received grants from Merck and Roche to supplement investigator-initiated studies that he leads at McGill University.

E.L.F. is Editor-in-Chief at Preventive Medicine and K.D.V. is an Assistant Editor at Preventive Medicine. The process of soliciting the special issue, sending out manuscripts for review, the peer-review process and editorial decision making was conducted entirely outside of the Preventive Medicine online system (for which Dr. Franco and Ms. Volesky have access to through their regular Preventive Medicine duties).

Funding sources: This research is supported by the Canadian Cancer Society Partner Prevention Research Grant (grant #703106).

Role of funding source: The sponsor had no role in the study design, data collection, analysis or interpretation of the data, writing of the manuscript or in the decision to submit this manuscript for publication.

HIGHLIGHTS

- Infections such as human papillomavirus (HPV) and hepatitis C virus cause cancer.
- Most cancer-causing infections can be prevented or treated.
- Infections were responsible for an estimated 3.7% (7097 cases) of new cancers in Canada in 2015.
- HPV was the leading infectious cause of cancer in Canada.

ABSTRACT

Infections are estimated to cause approximately 15% of the world's cancers with large geographic variations. Yet, Canadian estimates for specific cancer-causing infections are not available. To estimate the number of infection-associated cancers diagnosed among Canadian adults in 2015, we calculated population attributable risks (PARs) and the number of attributable cases for seven carcinogenic infections and their 20 associated cancers. A systematic literature search was performed for each infection to obtain data on infection prevalence in the population and the relative risk or odds ratio associated with the cancer it causes. When mechanistic evidence suggested that detection of a given infection within cancer tissue was sufficient to attribute the cancer to the infection, prevalence among cancer cases was used to approximate the PAR. Data from 61 studies formed the basis of our analyses. The estimated number of infection-attributable cancer cases for 2015 was: 3828 for human papillomavirus (HPV), 2052 for *Helicobacter pylori*, 578 for Epstein-Barr virus, 509 for hepatitis B and C viruses (HBV, HCV), 100 for human herpesvirus type 8, and 30 cases for human T-cell lymphotropic virus type 1. These seven infections were responsible for 3.7% of cancers diagnosed among Canadian adults in 2015; 3.5% among men and 4.0% among women. The infections with the highest number of attributable cases are largely preventable or treatable through vaccination (HBV and HPV), antibiotic therapy (*H. pylori*), or a combination of interventions (HCV), thereby representing an important target for reducing the infection-caused cancer burden among Canadians.

1. INTRODUCTION

Numerous infectious viruses and bacteria are established risk factors for certain cancers.[1] Many carcinogenic infections are strongly associated with specific cancers (i.e., *Helicobacter pylori* (*H. pylori)* and non-cardia gastric cancer, hepatitis B virus (HBV) and hepatitis C virus (HCV) and hepatocellular carcinoma),[42,77] while several others are necessary causes for cancer development (i.e., human papillomavirus (HPV) and cervical cancer, human herpes virus type 8 (HHV-8) and Kaposi sarcoma).[4,78]

Globally, almost one-sixth of cancers were attributable to infections with large geographical variations observed.[54-56] The proportion of infection-attributable cancers in 2012 varied from a high 31.3% in Sub-Saharan Africa to a low 4.0% in North America.[56] Although the latter constitutes a relatively smaller percentage, there is an opportunity to lower the Canadian cancer burden with currently available interventions. Specifically, primary preventive interventions include vaccination against HBV and HPV, along with secondary prevention measures such as direct-acting antivirals for chronic HCV infection and antibiotic therapy to treat *H. pylori* infection.[79-81] The prolonged latency associated with HCV and *H. pylori* provides an opportunity to treat them prior to cancer development.[82]

Although, to date, no study has estimated the impact of the different infections on cancer incidence in Canada, a global study reported that 3.9% of incident cancers in Canada were attributable to infections overall in 2012.[56] The global analysis combined infection prevalence for regions comprising many countries; for example, low, medium and high infection incidence areas. Since infection prevalence varies geographically, region-specific data based on more recent evidence from the scientific literature and population-based studies are necessary to obtain accurate estimates of the impact of infections on cancer incidence. Additionally, estimating individually the proportion of cancers attributable to each infection provides essential assessment of the cancer burden due to infections with modifiable prevalence.

Estimates of the impact of each infection on cancer incidence will contribute to the evidence needed to prioritize strategies aimed at reducing the prevalence of certain carcinogenic infections and initiating treatment for others. We estimated, among individuals 18 and older, the proportion and number of cancers diagnosed in Canada in 2015 that were attributable to infections, by sex and age whenever possible.

2. METHODS

The current analysis is part of the ComPARe (Canadian population attributable risk of cancer) Study, which estimates the current and future burden of cancer due to modifiable risk factors in Canada. Here, we estimated the current burden of cancers caused by infections.

2.1. Infections and cancer sites selection

We considered infections classified by the International Agency for Research on Cancer (IARC) as established, Group 1, carcinogens (**Table 1**). Infections with extremely low, prevalence in Canada (*Opisthorchis viverrini*, *Clonorchis sinensis*, and *Schistosoma haematobium*) were excluded. We also did not include human immunodeficiency virus (HIV) because HIV acts indirectly through immunosuppression, thereby amplifying the carcinogenic effects of co-infections such as Epstein-Barr virus (EBV), HCV, and HPV, infections that are already included in our analysis. Table 1 also enumerates the cancer sites for which there was 'sufficient' evidence for the role of infections in carcinogenesis, as concluded by IARC.[1] There was one exception; we estimated the impact of HPV16 on laryngeal cancer incidence because more data have accumulated since the last IARC monograph publication on HPV in support of an etiologic role of HPV in laryngeal cancer.[83,84]

2.2. Population attributable risk calculations

To estimate the proportion of cancer incidence that could have been avoided had the infection been eliminated, we calculated population attributable risks (PARs). The three equations below can estimate PARs for binary exposures (infected or not). The first formula requires the prevalence of the infection in the general population (Pe) and its relative risk (RR) or odds ratio (OR) associated with the cancer.[69] When Pe is not known, the second formula can estimate PARs using prevalence in cases (Pc) in place of Pe.[72] The third method is used when the attributable risk in the exposed approaches 1.0 (i.e., RRs are high, ~10+), such that the prevalence in cases approximates the PAR.

$$1.\ PAR = \frac{Pe(RR-1)}{1+Pe(RR-1)} \qquad 2.\ PAR = Pc\frac{(RR-1)}{RR} \qquad 3.\ PAR = Pc$$

Since we were able to obtain population-based data for HBV, HCV, and *H. pylori* prevalence, the first formula was used for estimating PARs for HBV, HCV, and *H. pylori*. The PARs for the remaining infections, EBV, HPV, HHV-8 and human T-cell lymphotropic virus type 1 (HTLV-1) were estimated with the third formula because they either demonstrate strong relationships with their associated cancers or mechanistic evidence exists for the role of the infection in cancer thus allowing for the PAR to be

approximated by the prevalence in cancer cases.[1,47,56]

2.3. Data collection and selection

The data needed to estimate PARs were identified by reviewing IARC monographs,[1,85,86] PAR analyses from other regions,[54,56,58,67] the Catalan Institute of Oncology HPV Information Centre reports for Canada and the United States,[87,88] and results of our systematic literature reviews. A systematic literature search was conducted for each infection (details in **Supplementary Table 1**) to extract data on the prevalence of the infection and identify meta-analyses on infection-associated cancer sites. Since the most recent IARC meeting on infectious agents considered data published to the end of 2007, we searched for records published in English or French from January 1, 2008 to the search date of June 20, 2017. When data were sparse, we performed more targeted searches in PubMed and contacted experts in their respective fields. Ethics approval was granted for this project by the Health Research Ethics Board of Alberta - Cancer Committee (HREBA.CC-14-0220_REN4), and McGill University exempted this study from Research Ethics Board review.

Cancers for which the infection is a necessary cause or part of the diagnostic criteria for a given cancer were: cervical cancer, extranodal natural killer T-cell lymphoma - nasal type, Kaposi sarcoma, primary effusion lymphoma, and adult T-cell leukemia/lymphoma, 100% were attributable to their associated infection and therefore inclusion criteria were not required. For all other infections and cancers, the inclusion criteria were: adult population (defined as age 15 and older), North American study population, non-specialized population (e.g. studies performed in exclusively HIV-positive participants were excluded), 10 or more cancer cases, and use of the gold standard method to detect the infection. The inclusion criteria specific to each infection-cancer pair are noted in the tables of included studies (**Supplementary Tables 2–13**).

When the prevalence in cancer cases approximated the PAR (formula 3), the infection had to be detected in the cancer tumor, such as in a biopsy or surgical specimen. To extrapolate prevalence estimates to recent cancer incidence, rather than incorporating a latency period, the aim was to select studies conducted closer to the timeframe when cancer incidence data were collected. For this reason, studies had to be published in 1995 or later. Specifically, the prevalence of any HPV in the oropharynx has increased over time in the United States; pre-1990 HPV prevalence was 20.9% and from 2000–2013 it rose to 65.4%,[89] further emphasizing the importance of utilizing more recent studies.

The prevalence of HBV and *H. pylori* were derived from North American population-based serosurveys, and HCV prevalence came from a study that modeled chronic HCV prevalence in the Canadian population.[90] Due to limited data, for the measures of association for *H. pylori* associated cancers, *a posteriori* decision was made to consider studies conducted among European populations and studies that used a detection method that preceded the current gold standard method (we corrected to the new standard).

The chosen detection method for assessing the presence of infection was crucial to the PAR estimation. Selecting studies that utilized the gold standard detection method was prioritized over other factors such as having a Canadian population or sex and age-specific results leading to sparser data.

2.4. Estimating infection prevalence in the Canadian population

Below is a brief description of how we adjusted population-based data to obtain sex- and age-specific estimates of HBV, HCV, and *H. pylori* prevalence for the Canadian population. The prevalence estimates and further details are provided in **Supplementary Tables S2–S5**.

2.4.1. Hepatitis B virus

The Canadian Health Measures Survey (CHMS) was the first population-based survey to provide estimates of HBV and HCV prevalence for the Canadian population.[91] Data from two cycles of the CHMS,[92] collected from 2007–2009 and 2009–2011, were combined for analysis. The combined participation rate for those providing direct health measures after sample strategy adjustments was 52.8% for the two cycles.[91] Sera from CHMS participants aged 14–79 testing positive for hepatitis B core antigen (anti-HBc) were then tested for hepatitis B surface antigen (HBsAg). Chronic HBV infection is defined as the presence of HBsAg six months after a positive HBV test.[93] Given the cross-sectional design of the CHMS, we assumed that HBsAg positivity at one time point represented chronic HBV infection. Privacy restrictions limited HBsAg results to either sex or broad age groups (14–49 and 50–79), yet sex and age effect HBV prevalence. To obtain Canadian age-specific prevalence estimates, we used the HBsAg 10-year age-group prevalence from two merged cycles of the weighted National Health and Nutrition Examination Survey (NHANES)[94,95] to partition the CHMS estimates by 10-year age groups. The first two cycles of the CHMS were collected from 2007 to 2011, resulting in a six-year latency. This time period does not correspond to the prolonged latency for hepatocellular carcinoma,[96] yet it is still plausible as the CHMS measured prevalent not incident HBV infection.

2.4.2 Hepatitis C virus

The CHMS is a household-based survey of non-institutionalized populations.[97] Thus, groups with higher HCV prevalence, namely intravenous drug users, were underrepresented by excluding those who were homeless or in prison. Moreover, although a diagnosis of either HBV or HCV in Canada are reported to national public health agencies,[98] many of these infections remain undiagnosed and therefore uncaptured in this data source. We thus obtained the modeled chronic HCV prevalence by birth cohort from Trubnikov, Yan and Archibald who accounted for high-risk groups and undiagnosed infections in their analyses.[90] To obtain chronic HCV prevalence by sex, we partitioned the estimates using the sex distribution of HCV prevalence from another study that modeled HCV prevalence in Canada in 2007.[99] Since the latency period between initial HCV infection and hepatocellular carcinoma is 25–30 years,[82] we used the midpoint of a 15-year latency in our estimates.

We did not estimate a PAR for HBV and HCV coinfection and hepatocellular carcinoma because data on coinfection prevalence were not available. To estimate the combined impact of HBV and HCV on hepatocellular carcinoma, we combined their PARs with the following equation: 1 – (1-PAR_HBV) * (1-PAR_HCV).[72]

2.4.3. Helicobacter pylori

Few studies have assessed *H. pylori* prevalence in Canadian populations. Although most of these studies were conducted with specialized populations,[100,101] one study included 1,306 residents aged 50–80 in Canada's most populous province, Ontario.[102] As population-based data covering a broad age range were required, we opted to utilize other data. *H. pylori* sera-status was assessed in one NHANES cycle collected from 1999–2000[103] which resulted in a 15–16 year latency period. The weighted NHANES data were reweighted by sex, five-year age groups, and race/ethnicity (Black, Latin American, White, and Other) to better reflect the composition of the Canadian population in 2001 (the closest year for which Canadian census ethnicity data were available). Missing *H. pylori* results, which made up 5.0–6.6% of the reweighted data, were assumed to be missing completely at random and excluded. Additionally, half of the 1–2% 'equivocal' results, which were the results for IgG level between the cut-offs for positive and negative results, were re-assigned as positive or negative. NHANES used enzyme-linked immunosorbent assay (ELISA) to detect *H. pylori*. ELISA has a sensitivity of 95.6% and specificity of 92.6%.[104] We corrected our reweighted prevalence data according to these reported diagnostic accuracy measures.[105]

Since immunoblot is more sensitive than ELISA for the detection of *H. pylori* in gastric cancer cases,[106,107] we also corrected the association measures from matched case-control studies that used ELISA by deriving a formula used to adjust the OR,[105] and calculating sensitivity and specificity parameters. The latter were derived by pooling the sensitivity and specificity from three studies,[106-108] that directly compared ELISA and immunoblot in the same patients.

2.5. Estimating infection prevalence in cancer cases

The PARs for EBV- and HPV- associated cancer sites were approximated by pooling studies that provided data on the prevalence of the infectious agent as detected in cancer tissues. For anogenital cancers, we considered an infection with at least one high-risk type (HPVs 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73, and 97) to indicate that the cancer was due to HPV. Head and neck cancers were considered attributable to HPV if genotype 16 was found via the detection of E6 and/or E7 oncoproteins which indicates viral activity and replication.

2.6. Cancer incidence

To determine the number of cases attributable to a given infection, the calculated PAR is multiplied by the number of incident cases. Incident cancer data were obtained from the Canadian Cancer Registry for 2015, which was the most recent year available. When data were requested for rare or subsite cancers, they were aggregated to maintain privacy; for example, cancer incidence counts were combined into two age groups (ages <50, and ≥50), instead of five-year age groups. To preserve the granularity in the incidence data, we estimated the proportion of liver cancer estimated to be hepatocellular carcinoma. A study using SEER (Surveillance, Epidemiology, and End Results) data reported that there were 55,344 primary liver cancers diagnosed from 1978–2007, of which 44,080 were hepatocellular carcinoma.[109] We applied the ensuing proportion of 0.797 (44,080/55,344) to liver cancer incidence to get the estimated number of hepatocellular carcinoma cases.

For the province of Quebec, the most recent year that cancer incidence data were available was 2010. Quebec's 2015 cancer incidence was estimated in one of two ways. For cancers with fewer cases (< 500 in Canada in 2015), the last five years of available incidence data for Quebec, 2005–2010, were averaged and applied to Quebec's 2015 population. For other cancers, Quebec's 2015 incidence was imputed by fitting a Poisson regression on Canada's 2008–2015 incidence.

2.7. Statistical analysis

To obtain the prevalence of a given infection in its associated cancer, individual studies were pooled with a random effects model; a fixed effect model was adopted if the index of consistency (I^2) was <25% and not statistically significant (p>0.05). To pool the proportions and measures of association, we used the commands *metaprop*[110] and *metan*,[111] respectively. To calculate 95% confidence intervals (CIs) for the pooled proportions, the exact method was used with the command "cimethod (exact)". When studies were excluded by the software because of inadmissible 95% CIs (e.g., proportions of 1.0 can yield CIs over 1.0), the Freeman-Tukey double arcsine transformation was enabled to calculate admissible 95% CIs bounded by 0.0–1.0 (Stata command is: "ftt"). All meta-analyses were conducted in Stata v14 (StataCorp., College Station, TX, USA). R was used to calculate PARs via formula 1.[112] For infections where the PAR was approximated by the prevalence of the infection in cancer cases, no additional calculations were necessary after pooling the prevalence. The CIs for PARs calculated via formula 1 were calculated as previously described.[113,114]

3. RESULTS

A summary of the overall methods and findings for HBV, HCV, *H. pylori* is presented in **Table 2**, and for infections where the prevalence in cases approximated the PAR in **Table 3**. Specific results and tabulations on the characteristics of included studies as well as forest plots, are provided under the respective infection and cancer sites (**Supplementary Tables 6–13** and **Figs. 1–8**).

Table 2 shows that the prevalence of chronic HBV infection in the Canadian population was <1.0% across all age and sex groups whereas chronic HCV prevalence ranged from 0.1 to 1.9%. The prevalence of *H. pylori* was notably lower among those younger than 50 years (12.8% for men and 9.8% for women) compared to those aged 50 years and over (27.9% for men and 29.6% for women). Between 1.6 and 15.3% of hepatocellular carcinomas were attributable to chronic HBV infection (**Supplement, Table S2**). Chronic HCV had higher attributable percentages than HBV, ranging from 2.5 to 30.0% (Supplement, Table S4). However, the percent of non-Hodgkin lymphoma attributable to HCV was negligible (<0.7%) for each age and sex group.

As shown in **Table 3**, the proportion of cancer attributable to high-risk HPV types in anogenital cancers was lowest for penile cancer (39.3%) and highest for cervical cancer (100.0%). The presence of HPV16 in head and neck cancers was 60.2% for the oropharynx, 12.7% for the larynx and 8.2% for the oral cavity.

Table 4 demonstrates that HPV infections were the causative agent for most infection-associated cancers (3,828, 95% CI: 3,190–4,425), followed by *H. pylori* (2,052, CI: 1,473–2,395), and EBV (579, CI: 286–604). More than half (54.0%) of the infection-caused cancers diagnosed in 2015 were related to HPV, then *H. pylori* (28.9%), EBV (8.1%), HBV/HCV (7.2%), HHV-8 (1.4%), finally HTLV-1 (0.4%) (data not shown). The cancers with the highest number of attributable cases were: non-cardia stomach (n = 1,730), cervix (n = 1,375), oropharynx (n = 1,083), anus (n = 589), and hepatocellular carcinoma (n = 480) (Table 4). A total of 7,097 cancers were attributable to infections, representing an estimated 3.7% of the cancers diagnosed among those ≥18 years old in 2015. The proportion of incident cancers attributable to infections was higher among women (4.0%) than men (3.5%).

4. DISCUSSION

The proportion of attributable cancers in Canada in 2015 ranged from a low of 0.4% for HCV in non-Hodgkin lymphoma to a high of 100.0% for HPV in cervical cancer. Cervical cancer was one of five cancers sites where all cases are attributable to an infection. With few exceptions (HCV in non-Hodgkin lymphoma, and HPV in the oral cavity and larynx), all the calculated PARs exceeded 25.0%, thereby demonstrating the important role that infections play in certain malignancies.

We found that the burden of infection-caused cancers was higher among women (4.0%) than men (3.5%), largely because of HPV's role in cervical and other anogenital cancers. Estimates for the United Kingdom also demonstrated a higher attributable proportion among women than men (3.7% versus 2.5%, respectively) in 2011[67] and a similar finding was found in Australia where 2.4% of cancers diagnosed among men in 2010 were attributed to infections and 3.7% among women.[58] In contrast, an analysis for the USA found that 3.3% among both men and women were attributable to infections in 2014.[17]

As PAR estimates assume causality between the exposure and outcome, we included only established carcinogens and cancer sites where the evidence for the role of the infection was deemed 'sufficient' by the IARC (except for HPV16 in laryngeal cancer). Yet, there is increasing evidence that other infection cancer associations including EBV in gastric carcinoma, HBV in non-Hodgkin lymphoma and HCV in cholangiocarcinoma, among others, may also cause cancer. If these associations were included, the impact of infections on cancer incidence would be higher than what we reported here.

4.1. Hepatitis B and C viruses, and H. pylori

The combined impact of the hepatitis viruses resulted in 27.4% of hepatocellular carcinoma incidence being attributable to HBV/HCV. Since HBV can be avoided with vaccination that began in Canada in the early 1980s, and because HCV can be prevented through a variety of behavioral interventions and treated with direct-acting antivirals, the future burden of hepatocellular carcinoma has the potential to decrease by reducing the prevalence of these viruses.

Globally, *H. pylori* was responsible for 89% of non-cardia gastric cancers.[46] We calculated that 68.8% of incident non-cardia gastric cancers in Canada were due to this infection. We estimated PARs based on elimination of the infection. This information is helpful for understanding the impact of infections on cancer incidence; however, in practice, elimination may not be entirely feasible. For example, *H. pylori* can be treated with quadruple antibiotic therapy, but challenges in the scalability of screening for the infection and concerns over antibiotic resistance limit the prospect of eliminating the infection at the population level.[7,115,116]

4.2. EBV, HHV-8 and HTLV-1

Although EBV is the infection with the highest prevalence with >90% of adults infected,[33] it was responsible for only 8.1% of the infection-caused cancers in Canada in 2015. In a similar vein, some infections with PARs of 100% were responsible for a small number of cancers (e.g. HHV-8 and HTLV-1) because of the rarity of cancers they cause.

4.3. Human papillomavirus

We found that 54% of the infection-associated cancers were due to HPV. This percentage is higher than the reported 29.5% global contribution of HPV to infection-associated cancers.[56] In particular our estimates for HPV16's role in head and neck cancers were higher than global estimates. Meta-analyses have reported higher HPV prevalence in oropharyngeal cancers in North American populations compared to other continents.[117,118] Our estimate of 60.2% with E6/E7 detection, albeit numerically similar to that of Ndiaye et al. (60.4%),[117] is actually higher than the latter because it represents detection of HPV16, whereas the 60.4% estimate in that study is for all HPV types combined. The oropharynx had the third highest number of attributable cases. Since 1997, oropharyngeal and oral cancer incidence has increased in Canada, especially among men, this is in part due to HPV's role in head and neck cancers.[119] The Canadian Cancer Society estimated that in 2012, cervical and oropharyngeal cancers each accounted for 35% of the HPV-associated cancer burden. We

too, found that approximately one-third of the HPV associated cancer burden was due to cervical (35.9%) and oropharyngeal (28.3%) cancers.[119] Since we examined the contribution of HPV16, any of the three available HPV vaccines provide coverage against this HPV type. Although a smaller proportion of oral cavity and laryngeal cancers are attributable to HPV (8.2% and 12.7%, respectively), they added 269 cases to the infection-associated cancer burden. School-based HPV immunization programs began in Canada in 2007. More recently, these programs have been extended to boys.[8] We found that one-third (34.0%) of HPV associated cancers were diagnosed among men, which further emphasizes the importance of vaccinating boys.

4.4. Limitations

The main limitation of our study was the lack of Canadian-specific infection data and the subsequent reliance on data collected in the United States and for *H. pylori* data collected from European populations. We have assumed that the exposure prevalence and strength of the relationship between the infection and cancer as observed in American and European populations were comparable to what would have been observed in Canada. For example, we reweighted the age, sex and race/ethnicity distribution from a population-based survey of *H. pylori* prevalence in the United States (NHANES) to match that of the Canadian population in the closest available year. Reweighting assumed that differences in the prevalence of *H. pylori* between the two countries were due to age, sex, and race/ethnicity – but these variables do not likely fully account for the potential differences between Canada and the United States. For some infection cancer site pairs, irrespective of including data collected outside of Canada and performing more targeted literature searches, the data remained sparse. This situation was particularly true for: *H. pylori* and gastric mucosa-associated lymphoid tissue lymphoma, EBV and Burkitt lymphoma, and HPV and vaginal cancer. This result was anticipated since the cancer sites with sparser evidence were also the rarer cancers.

Focusing exclusively on Canada allowed us to obtain much of the rare and subsite cancer incidence data we required for accurate estimates of the number of attributable cases. However, we estimated rather than directly obtained hepatocellular carcinoma and Quebec's cancer incidence. For cancer sites with fewer than 500 cases in Canada in 2015, the five-year incidence rates were averaged but this averaging relies on assumptions that the average of the last five years of available cancer incidence for Quebec (2006–2010) is representative of the 2015 cancer incidence and that the trend has remained stable.

Since we used existing data, our findings inherited the methodologic flaws of included studies and population-based surveys. This concern was at least partially mitigated by including only those studies that met stringent inclusion criteria aimed at enhancing the validity of our estimates. We attempted to correct for measurement error; however, some error may remain. Additionally, our correction for error in assessing the association between *H. pylori* and non-cardia gastric cancer did not account for confounders. Although the included studies were matched case-control studies, unmatched confounders are then not adjusted for.

By not conducting a separate analysis for HIV-1, we potentially underestimated the impact of infections on cancer incidence. The proportion of cancer attributable to EBV has the potential to increase since non-Hodgkin lymphomas among HIV positive populations were not included in this analysis.

5. CONCLUSION

We estimated that 3.7% of cancers diagnosed among Canadians aged 18 and older in 2015 were attributable to seven carcinogenic infections. This percentage translated into 7,097 cancers, where ~6,400 could potentially be prevented with currently available vaccines or treatments. HPV was responsible for more cancers than other infection, comprising more than half of the infection-associated cancer burden. The presence of three vaccines that confer 95% efficacy against the HPV types responsible for cancer incidence is encouraging. Although Canada has a lower infection-associated cancer burden relative to many other countries,[56] infection-associated cancers continue to impact cancer incidence and increasing vaccine hesitancy has the potential to limit the progress that could be made in reducing the HPV and HBV associated cancer burden.

Table 1. Overview of the carcinogenic infections and associated cancer sites[a]

Infection	Main transmission route(s)	Risk factors for transmission	Carcinogenic mechanism(s)[b] *From Bouvard 2009*	Gold standard for detection	Cancers with *sufficient* evidence[c]	Cancers with *limited* evidence[c]
Hepatitis B virus (HBV), *chronic infection*	Sera and other body fluids	Reusing needles, sexual intercourse	Inflammation Liver cirrhosis Chronic hepatitis	HBsAg	Hepatocellular carcinoma	Cholangiocarcinoma Non-Hodgkin lymphoma
Hepatitis C virus (HCV), *chronic infection*	Sera	Reusing needles	Inflammation Liver cirrhosis Liver fibrosis	HCV RNA	Hepatocellular carcinoma, non-Hodgkin lymphoma	Cholangiocarcinoma
***Helicobacter pylori* (*H. pylori*)**	Oral/fecal	Crowding, contaminated water	Inflammation Oxidative stress Altered cellular turn-over and gene expression Methylation Mutation	Immunoblot	Non-cardia gastric carcinoma, low-grade B-cell MALT gastric lymphoma	None
Epstein-Barr virus[d] (EBV)	Oral/saliva	Pre-chewing food for babies, sharing utensils, kissing [(1)]	Cell proliferation Inhibition of apoptosis Genomic instability Cell migration	EBER ISH LMP1 IHC for Hodgkin lymphoma[(120)]	Burkitt lymphoma Hodgkin lymphoma, extranodal natural killer T-cell lymphoma - nasal type, nasopharyngeal carcinoma, immune suppression-related non-Hodgkin lymphoma	Gastric carcinoma Lymphoepithelioma-like carcinoma
Human papillomavirus (HPV), type 16	Skin-to-skin/ mucosal	Sexual contact including oral sex and open mouth kissing	Immortalisation, Genomic instability Inhibition of DNA damage response Anti-apoptotic activity	PCR alone or with p16 for anogenital cancers E6 and/or E7 mRNA for head and neck cancers	Cancers of the cervix, anus, penis, vagina, vulva, oropharynx, tonsil, and oral cavity	Laryngeal carcinoma
Human herpesvirus, type 8[e] (HHV-8)	Oral/saliva	Sexual contact including oral sex and open mouth kissing	Cell proliferation, Inhibition of apoptosis Genomic instability Cell migration	IFA	Kaposi sarcoma, primary effusion lymphoma	Multicentric Castleman's disease
Human T-cell lymphotropic virus, type 1 (HTLV-1)	Sera and other body fluids, including breast milk	Breast-feeding, sexual intercourse, and reusing needles [(121)]	Immortalisation and transformation of T cells	PCR	Adult T-cell leukemia/lymphoma	None

HBsAg = Hepatitis B surface antigen, RNA = ribonucleic acid, mRNA = messenger ribonucleic acid, EBER ISH = Epstein-Barr virus encoding region *in situ* hybridization, LMP1 = latent membrane protein 1, IHC = immunohistochemistry, PCR = polymerase chain reaction, IFA = immunofluorescent assays, MALT = mucosa associated lymphoid tissue

a. Included infections have been categorized by IARC as Group 1 carcinogens.
b. Carcinogenic mechanisms were taken from Bouvard 2009.[(2)]
c. Cancer sites were categorized by IARC are having *sufficient* or *limited* evidence.
d. Epstein-Barr virus is also referred to as Human herpes virus, type 4.
e. Human herpesvirus, type 8 is also referred to as Kaposi sarcoma virus.

Table 2. Infections for which the attributable risk was estimated using the prevalence in the population and measures of association

Infection Cancer (ICD-03 code)	**Method of infection measurement**	**Source of prevalence data**	**Range of prevalence estimates by sex**	**Data used to estimate measure of association**	**Odds ratio (95% CI)**
Helicobacter pylori					
Stomach, non-cardia (C16.1–16.9)	Serology with ELISA or immunoblot detection	NHANES (1999–2000) data reweighted by Canada's sex, age, and ethnicity distribution. Estimates were corrected for sensitivity and specificity.	Men 12.8% (aged <50) to 27.9% (aged ≥50) Women 9.8% (aged <50) to 29.6% (aged ≥50)	Pooled unadjusted ORs from matched case-control studies with fixed effects: 3 corrected studies that used ELISA and 3 studies that used immunoblot	9.4 (6.5–13.4)
Stomach, MALT lymphoma (9699)	Serology with ELISA detection			1 study of 33 cases matched to 134 controls [122]	6.3 (2.0–19.9)
Hepatitis B virus					
Hepatocellular carcinoma (C22.0, 817)	Serology with HBsAg	CHMS HBsAg data (2007–2011) partitioned with NHANES HBsAg 10-year age group distribution (2007–2010)	Men 0.1% (aged 70–79) to 0.9% (aged 30–39) Women 0.1% (aged 70–79) to 0.7% (aged 30–39)	Meta-analysis with pooled estimate from 3 case-control studies conducted in the USA and 1 cohort study from Australia [42]	20.3 (11.3–36.5)
Hepatitis C virus					
Hepatocellular carcinoma (C22.0, 817)	Estimates from modelling studies [90,99]	Chronic HCV prevalence modeled for the Canadian population by 5-year birth cohorts,[90] partitioned with the sex distribution from another modelling study [99]	Men 0.2% (aged 16–20) to 1.9% (aged 46–50) Women 0.1% (aged 16–20) to 1.2% (aged 46–50)	Pooled from 7 studies from the USA and Australia [42]	23.8 (16.9–33.5)
Non-Hodgkin lymphoma (9591)				Adjusted OR calculated from SEER data with 33,940 cases matched to controls on sex, age, and year of diagnosis [123]	1.35 (1.06–1.73)

CI = confidence interval, MALT = mucosa-associated lymphoid tissue, NHANES = National Health and Nutrition Examination Survey, CHMS = Canadian Health Measures Survey, HBsAg = Hepatitis B surface antigen, SEER = Surveillance, Epidemiology, and End Results-Medicare (United States)

Table 3. Methods used for the infections where population attributable risks were estimated using the prevalence of infection in cancer cases

Infection Cancer (ICD-O-3 code)	Method of infection measurement	Source of prevalence estimates[d]	Cases used to estimate PAR, N	Sex/age group	PAR (Prevalence of infection in cancer cases)	
					Estimate (%)	95% CI
Epstein-Barr virus						
Burkitt lymphoma (9687)	EBER ISH	1 study	30	<50 years old	40.0	22.7–59.4
			21	≥50 years old	28.6	11.3–52.2
ENKTL, nasal type (9719)		Part of diagnostic criteria	--	All	100.0	--
Hodgkin lymphoma (C81)	EBER ISH and/or LMP1 IHC	4 studies	560	Men	43.0	28.4–57.7
			583	Women	26.6	12.1–41.1
Nasopharynx (C11)	EBER ISH	2 studies	172	All	69.4	61.9–76.9
Human papillomavirus, high-risk types,[a] anogenital tract cancers						
Anus (C21)	PCR detection with genotyping of at least HPV 16 and 18	5 studies	154	Men	87.6	76.4–95.8
			250	Women	94.6	89.3–98.3
Cervix (C53)		Necessary cause	--	Women	100.0	--
Penis (C60)		6 studies	311	Men	39.3	21.8–56.9
Vagina (C52)		2 studies	85	Women	72.2	62.7–81.7
Vulva (C51)		2 studies	43	<50 years old	76.8	64.2–89.4
		3 studies	201	≥50 years old	43.2	13.9–72.5
Human papillomavirus, type 16, head and neck cancers						
Oropharynx[b] (C01.9, C02.8, C02.4, C05.1, C05.2, C14.2, C09, C10)	PCR with E6 and/or E7 for HPV16	16 studies	1,396	All	60.2	51.8–68.5
Oral cavity[c] (C00.4–0.5, C00.9, C02.0–C02.9, C03, C04, C05.0, C05.8, C05.9, C06, C14.8)		9 studies	733	All	8.2	3.6–14.2
Larynx (C32)		5 studies	194	All	12.7	3.7–25.4
Human herpesvirus, type 8						
Kaposi sarcoma (9140)	IFA	Necessary cause	--	All	100.0	--
Primary effusion lymphoma (9678)	IFA	Part of diagnostic criteria	--	All	100.0	--
Human T-cell lymphotropic virus, type 1						
Adult T-cell leukemia and lymphoma (9827)	PCR	Necessary cause	--	All	100.0	--

EBER ISH = EBV-encoded RNA *in situ* hybridization, PCR = polymerase chain reaction, LMP1 = latent membrane protein 1, IHC = immunohistochemistry, CI = confidence interval, PAR = population attributable risk, ENKTL = extranodal natural killer T-cell lymphoma, IFA = immunofluorescent assays

-- CI not necessary.

a. High-risk HPV types include types classified by the International Agency for Research on Cancer as Group 1 (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 & 59), Group 2A (68) and Group 2B (34, 53, 66, 70 & 73) carcinogens. HPV type 97 was also considered a high-risk type.

b. Oropharynx subsites: base of the tongue (C01.9), overlapping lesion of tongue (C02.8), lingual tonsil (C02.4), soft palate (C05.1), uvula (C05.2), Waldeyer ring (C14.2), tonsil (C09), oropharynx (C10).

c. Oral cavity subsites: mucosa of lip (C00.4–0.5) and lip NOS (C00.9), other and unspecified parts of tongue (C02.0–C02.9), gum (C03), floor of mouth (C04), palate (hard, overlapping lesion, NOS) C05.0, C05.8, C05.9), other and unspecified parts of mouth (C06) and overlapping lesion of lip, oral cavity and pharynx (C14.8)

d. Included studies can be found in the supplement under their respective infection and cancer sites.

Table 4. Summary of the number of cases and proportion of cancers attributable to infections in Canada in 2015

Infection Cancer site(s)	Total			Men			Women		
	Obs cases[a]	AC[b]	% Attributable[c]	Obs cases	AC	% Attributable	Obs cases	AC	% Attributable
Hepatitis B and C virus									
Hepatocellular carcinoma	1750	480	27.4	1345	400	29.7	405	80	19.8
Hepatitis C virus									
Non-Hodgkin lymphoma	8290	29	0.4	4620	19	0.4	3670	10	0.3
Helicobacter pylori									
Stomach, MALT lymphoma	560	322	57.5	265	151	57.0	295	171	58.0
Stomach, non-cardia	2515	1730	68.8	1445	993	68.7	1,070	737	68.9
Epstein-Barr virus									
Burkitt lymphoma	85	30	35.3	65	23	35.4	20	7	35.0
ENKTL – nasal type	25	25	100.0	15	15	100.0	10	10	100.0
Hodgkin lymphoma	940	336	35.8	525	226	43.0	415	110	26.6
Nasopharynx	270	187	69.4	195	135	69.4	75	52	69.4
Human papillomavirus, high-risk types									
Anus	640	589	92.0	225	197	87.6	415	392	94.5
Cervix	1375	1375	100.0	…	…	…	1375	1375	100.0
Penis	205	81	39.3	205	81	39.3	…	…	…
Vagina	180	130	72.2	…	…	…	180	130	72.2
Vulva	635	301	47.4	…	…	…	635	301	47.4
Human papillomavirus, type 16									
Oropharynx[f]	1800	1083	60.2	1380	830	60.2	420	253	60.2
Oral cavity	1560	127	8.2	940	77	8.2	620	51	8.2
Larynx	1115	142	12.7	925	118	12.7	190	24	12.7
Human herpesvirus, type 8									
Kaposi sarcoma	90	90	100.0	70	70	100.0	20	20	100.0
Primary effusion lymphoma	10	10	100.0	10	10	100.0	…	…	100.0
Human T-cell lymphotropic virus, type 1									
Adult T-cell leukemia and lymphoma	30	30	100.0	15	15	100.0	15	15	100.0
All Associated Cancers[d]	**22,075**	**7097**	**32.2**	**12,245**	**3360**	**27.4**	**9830**	**3738**	**38.1**
All Cancers[e]	**189,530**	**7097**	**3.7**	**96,070**	**3360**	**3.5**	**93,460**	**3738**	**4.0**

Obs = observed, AC = attributable cases, MALT = mucosa-associated lymphoid tissue, ENKTL = extranodal natural killer T-cell lymphoma

a. Cancer incidence data for the year 2015 from the Canadian Cancer Registry. Quebec's cancer incidence was estimated. Hepatocellular carcinoma incidence was estimated by applying the proportion 0.79 to liver cancer incidence.

b. Number of cancer cases at individual cancer sites that can be attributed to infection.

c. Proportion attributable was calculated by dividing the number of cases attributable to infection by the number of the associated cancer cases. It differs from PAR which for some cancer sites varied by age and/or sex.

d. All associated cancers includes all cancers known to be associated with infections listed in the table.

e. All cancers includes all incident cancer cases in Canada among those 18 and older in 2015.

f. Includes the base of the tongue and tonsils.

Supplementary material to manuscript #1

Systematic literature searches

We performed two PubMed searches per infection (**Table S1**). The first search strategy aimed at identifying relevant meta-analyses to extract reported pooled estimates or to identify and meta-analyze individual studies by pooling pertinent results from the identified meta-analyses. The second strategy aimed at identifying infection prevalence data via medical subject heading (Mesh) terms. The search criteria were (1) studies published in English or French, (2) studies published from January 1, 2008 up to the search date of June 20, 2017 (the last International Agency for Research on Cancer monograph (volume 100B) examined the literature published to the end of 2007), and (3) The keywords "Not" and "genetics" were added to each search to increase the likelihood of finding relevant records.

Table S1. Systematic search strategies in PubMed and number of reviewed records

Infection	Strategy 1: Meta-analyses		Strategy 2: Prevalence (Mesh terms)		Total records reviewed, n
	Search terms	Unique records, n	Search terms	Unique records, n	
Hepatitis B virus	Hepatitis B virus OR HBV AND cancer AND meta-analysis	87	Neoplasms AND prevalence OR meta-analysis AND Hepatitis B	150	237
Hepatitis C virus	Hepatitis C virus OR HCV AND cancer AND meta-analysis	59	Neoplasms AND prevalence OR meta-analysis AND Hepatitis C	141	200
Helicobacter pylori	Helicobacter pylori OR H. pylori AND cancer AND meta-analysis	129	Neoplasms AND prevalence OR meta-analysis AND Helicobacter pylori	155	284
Epstein-Barr virus	Epstein Barr virus OR EBV AND cancer OR gastric cancer OR stomach cancer AND meta-analysis	46	Neoplasms AND prevalence OR meta-analysis AND Epstein-Barr Virus Infections OR Herpesvirus 4, Human	45	91
Human papillomavirus	Human papillomavirus OR HPV AND cancer AND meta-analysis	223	Neoplasms AND prevalence OR meta-analysis AND Papillomaviridae	305	528
Human herpesvirus, type 8	--	--	Neoplasms AND prevalence OR meta-analysis AND Sarcoma, Kaposi OR Herpesvirus 8, Human	53	53
Human T-cell lymphotropic virus, type 1	Human t cell lymphotropic virus type 1 OR HTLV-1 AND cancer AND meta-analysis	84	Neoplasms AND prevalence OR meta-analysis AND Human T-lymphotropic virus 1	21	105
Total		**628**		**870**	**1,498**

MeSH = medical subject headings

-- Indicates that we did not perform a search for meta-analyses, as HHV-8 is rare in Canada and a necessary cause of associated cancers.

HEPATITIS B VIRUS (HCV)

Less than 2% of the Canadian, American as well as Northern and Western European populations are infected with chronic HBV.[124] Progression from acute to chronic HBV is lower among adults (<5%) compared to infants (80–90%).[125]

Estimating chronic hepatitis B virus prevalence in Canada

The Canadian Health Measures Survey (CHMS) data contained sex but not age Hepatitis B surface antigen (HBsAg) prevalence. To incorporate age information into the CHMS sex estimates, we adjusted the data by the HBV prevalence distribution in the National Health and Nutrition Examination Survey (NHANES) from the United States. After excluding individuals under age 18, we merged two NHANES datasets collected from 2007–2008 and 2009–2010 to reflect the time-period covered by the CHMS cycles 1 and 2 collected from 2007–2009 and 2009–2011.[94,95] We partitioned the CHMS sex prevalence estimates by estimated HBsAg prevalence by 10-year age groups from the merged NHANES data. We chose to group results by a 10-year age group period because the 5-year age groups categorization resulted in having some groups with no positive results (e.g., sample size was too low to capture a positive result).

For each sex, the following process was carried out. Relative weights of HBsAg prevalence were assigned by dividing each NHANES age-group prevalence by the prevalence in the highest HBsAg prevalence group, aged 30–39 years. To obtain the Canadian weighted population for each age group, these relative weights were multiplied by the 5-year averaged Canadian population from 2007–2011. To obtain the adjusted Canadian prevalence of HBsAg for the reference group (aged 30–39), the 5-year averaged Canadian population was summed across all age groups and multiplied by the CHMS prevalence (0.54% for men and 0.36% for women), then divided by the sum of the weighted population for all ages. The adjusted Canadian prevalence for the other remaining age groups was calculated by multiplying the relative weight of the adjusted Canadian prevalence for ages 30–39. The estimated prevalence ranged from 0.12–0.93% (**Table S2**).

Hepatocellular carcinoma (HCC)

Through inflammation of the liver, known as cirrhosis, HBV can indirectly cause the major liver cancer histological type – HCC.[126] Our search provided one meta-analysis; Cho et al. 2011 reported a pooled OR of 20.3 (95% CI: 11.3–36.5) for the association between HBV and HCC for low HBV

prevalence areas which included three American studies and one Australian study.[42] The calculated PARs ranged from 2.3% among women aged ≥80 to 15.3% among men aged 40–49 (**Table S2**).

Table S2. Chronic hepatitis B virus prevalence estimates in the Canadian population and associated population attributable risks (%) for hepatocellular carcinoma, by age groups and sex[a]

Age group (in years)	NHANES chronic HBV prevalence, %		Adjusted Canadian chronic HBV prevalence, %				PARs for hepatocellular carcinoma, %			
	Men	Women	Men		Women		Men		Women	
			Prevalence	95% CI [b]	Prevalence	95% CI [b]	Estimate	95% CI	Estimate	95% CI
18–29	0.29	0.17	0.31	0.04–0.92	0.32	0.05–0.77	…	…	…	…
30–39	0.87	0.37	0.93	0.32–1.71	0.70	0.14–1.50	5.63	0.00–10.94	5.88	0.00–11.41
40–49	0.30	0.21	0.32	0.10–0.64	0.39	0.08–0.90	15.27	1.96–26.77	11.88	0.00–22.34
50–59	0.82	0.08	0.88	0.30–1.59	0.15	0.02–0.41	5.88	0.13–11.30	6.98	0.00–13.48
60–69	0.47	0.19	0.51	0.11–1.08	0.35	0.07–0.87	14.58	1.93–25.61	2.89	0.00–5.69
70–79	0.08	0.06	0.08	0.01–0.19	0.12	0.02–0.34	8.97	0.00–17.13	6.35	0.00–12.30
≥ 80	--	--	--	--	--	--	1.56	0.00–3.09	2.28	0.00–4.51

NHANES = National Health and Nutrition Examination Survey, HBV = hepatitis B virus, PAR = population attributable risk, CI = confidence interval

-- Indicates the lack of hepatitis serology data from individuals aged 80 and older in the Canadian Health Measures Survey (CHMS).

… After applying a latency, we could not estimate PARs for individuals aged 18–29.

a. Inclusion criteria: population-based serosurvey, measurement of HBsAg, adult study population (15 and older), age and sex data (or the ability to adjust for age and/or sex).

b. Monte Carlo simulated 95% CIs.

HEPATITIS C VIRUS (HCV)

Approximately 2% of people are infected with HCV globally.[96] Once infected with HCV the likelihood of progression to chronic infection is very high, at an average of 74% (95% CI: 71–78).[127] It was estimated that almost a quarter of a million Canadians (245,987) were living with chronic HCV infection in 2011.[90]

Estimating chronic hepatitis C virus prevalence in Canada

To include sex information in the chronic HCV prevalence estimates modelled for 5-year birth cohorts for the year 2000,[90] we partitioned using the HCV distribution from a study that modelled combined acute and chronic HCV infection prevalence in Canada for the year 2007.[99] We assumed that the combined acute and chronic HCV prevalence distribution by sex would be comparable to that of individual chronic infection due to the high proportion (~74%) of HCV infections progressing to chronic infection. As the two previously mentioned modelling studies estimated prevalence for the Canadian population,[90,99] census data for the relevant years were used to standardize the reported prevalence before partitioning the birth cohort estimates by sex. The estimated prevalence of chronic HCV infection was higher for men than women for all age groups (**Table S3**).

Table S3. Chronic hepatitis C virus prevalence estimates in the Canadian population (%), by age groups and sex[a]

Age group in 2000 (in years)[b]	Chronic prevalence in 2000,[c] %	Acute and chronic HCV prevalence in 2007,[d] %		Adjusted chronic prevalence in 2000, %	
		Men	Women	Men	Women
16–20	0.16	1.28	0.71	0.20	0.11
21–25	0.38	1.45	0.82	0.48	0.27
26–30	0.61	1.43	0.84	0.76	0.45
31–35	1.04	1.44	0.87	1.29	0.79
36–40	1.22	1.48	0.92	1.50	0.93
41–45	1.48	1.52	0.97	1.82	1.14
46–50	1.54	1.49	0.98	1.88	1.21
51–55	0.90	1.39	0.98	1.05	0.75
56–60	0.91	1.22	0.92	1.04	0.79
61–65	0.99	0.99	0.81	1.10	0.89
66–70	0.73	0.84	0.78	0.76	0.70
71–75	0.49	0.94	0.92	0.51	0.47
76–80	--	--	--	--	--
≥80	--	--	--	--	--

HCV = hepatitis C virus, PAR = population attributable risk

-- Indicates that the estimate was not available in the original study by Trubnikov, Yan & Archibald, 2014.

a. Inclusion criteria: population-based serosurvey, adult study population, no exclusions of groups where HCV is prevalent (e.g. those who are homeless or in prison).

b. Prevalence in 2000 was utilized to incorporate a 15-year latency period from chronic HCV infection to cancer diagnosis in 2015.

c. Prevalence was modelled using the back-calculation method as described by Trubnikov, Yan & Archibald, 2014.

d. HCV prevalence (acute and chronic infection), modelled by Remis 2010, provided the sex distribution of HCV that was used to adjust the birth cohort estimates from Trubnikov, Yan & Archibald, 2014.

Latency period for HCV

A 15-year latency was incorporated by applying the HCV prevalence estimates for the year 2000 to cancer incidence in the year 2015. For example, the prevalence of HCV in individuals aged 16–20 years in 2000 was applied to cancer incidence for those aged 30–34 years in 2015.

Hepatocellular carcinoma (HCC)

In the same way that HBV can inflame the liver, eventually leading to HCC, HCV follows a similar pathway.(126) The meta-analysis performed by Cho et al. provided a pooled estimate of the association between HCV and HCC (OR = 23.8, 95% CI: 16.9–33.5), based on six studies conducted in the USA and one from Australia.(42) PARs were higher for HCV compared to HBV for HCC (**Table S4**).

Non-Hodgkin lymphoma (NHL)

An OR of 1.35 (95% CI: 1.06–1.73) for the association between HCV and NHL was reported from a study with 33,940 NHL cases from the Unites States' Surveillance, Epidemiology, and End Results-Medicare database.(123) This modest measure of association coupled with low chronic HCV prevalence (<2.0%) resulted in PARs ranging from 0.04–0.65% (**Table S4**).

Table S4. Population attributable risks for chronic hepatitis C virus (%), by age groups and sex

Age at diagnosis (in years)[a,b]	PARs for hepatocellular carcinoma, %		PARs for non-Hodgkin lymphoma, %	
	Men	Women	Men	Women
30–34	4.44	2.53	0.07	0.04
35–39	9.90	5.86	0.17	0.10
40–44	14.80	9.38	0.27	0.16
45–49	22.71	15.22	0.45	0.27
50–54	25.52	17.54	0.52	0.33
55–59	29.34	20.62	0.63	0.40
60–64	29.95	21.61	0.65	0.42
65–69	19.39	14.54	0.37	0.26
70–74	19.12	15.19	0.36	0.27
75–76	20.00	16.84	0.38	0.31
80–84	14.76	13.81	0.27	0.25
≥85	10.36	9.77	0.18	0.17

PAR = population attributable risk

a. PAR estimates start at age 31 because a latency of 15 years was applied.

b. PAR estimates were applied to cancer incidence data for similar age groups; for example, age 31–35 was applied to incidence for those age 30–34.

HELICOBACTER PYLORI

H. pylori, a spiral shaped bacterium, was first identified in 1913 in the gastric mucosa.[128] Whether an *H. pylori* infection leads to cancer is influenced by bacterial virulence factors and host responses that themselves are influenced by factors, such as high salt concentration and low iron.[129] The global prevalence of *H. pylori* is about 50% and varies by socio-demographic and economic characteristics.[35]

Estimating *H. pylori* prevalence in the Canadian population

The one cycle of the National Health and Nutrition Examination Survey (NHANES) from the USA assessed *H. pylori* serostatus collected data via enzyme-linked immunosorbent assay (ELISA) from 1999–2000. These data were reweighted to match Canada's age, sex, and ethnic distribution. Next, the data were corrected for measurement error associated with ELISA. ELISA has a reported sensitivity (Se) of 95.6% and specificity (Sp) of 92.6%.[104] These parameters we used to adjust the uncorrected *H. pylori* prevalence (Pu) to the corrected prevalence (P) with the following equation:

$$P = \frac{(Pu + Sp - 1)}{(Se + Sp - 1)}$$

The overall prevalence of *H. pylori* prior to correction was 23.3% for men and 22.5% for women (**Table 2**). After correction, it was lower at 18.0% for men and 17.2% for women.

Table S5. Estimated prevalence of *H. pylori* in the Canadian population, by age groups and sex

Sex	Age-group (years)	*H. pylori* prevalence,[a,b] uncorrected		*H. pylori* prevalence,[c,d] corrected	
		%	95% CI	%	95% CI
Men	18–49	18.7	14.4–23.2	12.8	7.9–17.9
	≥50	32.0	26.3–38.1	27.9	21.4–34.8
	Total	23.3	19.8–26.9	18.0	14.1–22.1
Women	18–49	16.0	12.4–19.8	9.8	5.7–14.0
	≥50	22.5	19.4–25.8	29.6	23.2–36.3
	Total	22.5	19.4–25.8	17.2	13.6–20.8
Total	18–49	17.3	14.5–20.2	11.3	8.1–14.6
	≥50	32.8	28.8–37.0	28.8	24.2–33.6
	Total	22.9	20.5–25.3	17.6	14.9–20.3

a. Data from NHANES cycle 1999–2000.
b. IgG of ≥ 1.10 IgG was categorized as positive and < 0.90 IgG as negative. Equivocal samples (0.91–1.09 IgG) remained equivocal after repeat testing via another ELISA assay. Unequivocal samples were distributed 50/50 among positive and negative.
c. Data were reweighted to Canada's sex, age (5-year age group) and ethnicity (Black, Latin American, White, and Other) in the closest year available, 2001.
d. Missing results were excluded from the analysis.

Stomach cancer (non-cardia)

As the precursor to stomach cancer, stomach atrophy, progresses, the detectible bacterial load of *H. pylori* decreases.[46] This differential loss in sensitivity is mitigated by measuring *H. pylori* infection roughly 10 years before diagnosis, thereby making prospective data an important inclusion criterion. The finding that immunoblot is more sensitive than ELISA in detecting *H. pylori* necessitated a correction for this potential error.[46] The sensitivity and specificity were extracted and pooled from three studies that compared ELISA to immunoblot head-to-head.[106-108] A derivation of a formula used to correct measurement error (93% sensitivity and 83% specificity) in the ORs was applied to the three prospective case-controls that used ELISA (**Table S6**).[105] The corrected and immunoblot studies were pooled with fixed effects due to a lack of heterogeneity (**Fig. S1**).

Table S6. Characteristics of studies on the association between *H. pylori* and non-cardia gastric cancer

Study[a]	Study population	Mean or median follow-up years	Matching variables	Cases		Controls		Unadjusted OR (95% CI)	Corrected OR[b] (95% CI)
				n/N	Positive %	n/N	Positive %		
Studies that used ELISA or EIA to detect *H. pylori*									
Hansen 2007[130]	Norwegian cohorts Recruited: 1972–1986 Diagnosed: 1972–199	11.9	Sex, age, cohort, sera collection date	116/129	89.9	247/376	65.7	4.7 (2.5–9.4)	13.7 (5.5–34.4)
Knekt 2006[131]	Finnish cohort Recruited: 1968–72 Diagnosed: 1968–1991	Up to 24	Sex, age, municipality	176/193	91.2	292/372	78.5	2.8 (1.6–5.27)	8.9 (3.5–22.4)
Nomura 2002[132]	US cohort of men of Japanese ancestry Recruited: 1967–1977 Diagnosed: 1967–1996	12.7	Age, sera collection date	231/261	88.5	193/261	73.9	2.7 (1.66–4.50)	5.4 (3.0–9.8)
Studies that used immunoblot to detect *H. pylori*									
Gonzalez 2012[106]	10 European countries in the EPIC cohort Recruited: 1992–1998 Diagnosed: 2000–2004	10.7	Sex, age group, center and date of blood collection	82/88	93.2	199/338	58.9	9.6 (4.1–22.5)	--
Mitchell 2008[107]	Australian cohort Recruited: 1990–1994 Diagnosed: 1990–2002	11.6	Sex, age, birth country, sera collection date	32/34	94.1	85/134	63.4	9.2 (2.1–40.2)	--
Simán 2007[133]	Swedish cohort Recruited: 1974–1992 Diagnosed: –2000	Ranged from 9.2–12.6	Sex, age, and sera collection date	65/67	97.0	147/250	58.8	22.8 (5.5–95.1)	--

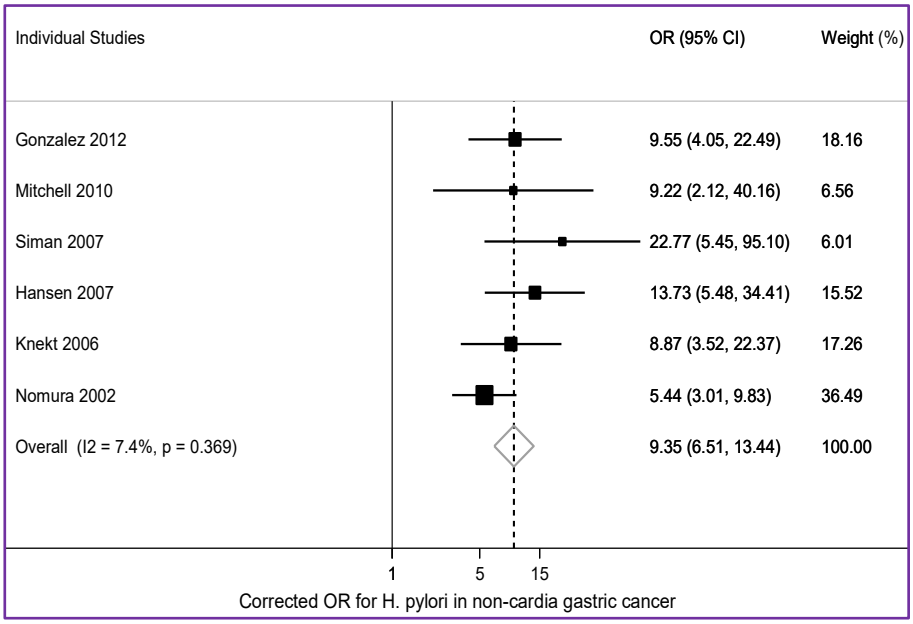
US = United States, CA = Canada, ELISA = enzyme-linked immunosorbent assay, EIA = enzyme immunosorbent assay, EPIC = European Prospective Investigation into Cancer and Nutrition

– Indicates that correction was not required.

a. Inclusion criteria: prospective serology collection (~10 years in advance of diagnosis), ELISA or immunoblot detection, 10 or more non-cardia gastric cancer cases, North American or European study populations, data required to correct sensitivity and specificity or immunoblot detection.

b. Corrected to 93% sensitivity and 83% specificity. ORs were calculated based on the condition maximum likelihood estimates, and CIs were based on Fisher exact tests.

Fig. S1. Forest plot of the association between H. pylori and non-cardia gastric cancer



OR = odds ratio

Gastric mucosa-associated lymphoid tissue (MALT) lymphoma

MALT lymphoma is most often diagnosed in the stomach, but can also be found in the lungs, thyroid, skin or soft tissues.[134] It is a type of non-Hodgkin lymphoma (NHL). The Canadian Cancer Society noted that the majority (60% or greater) of people with gastric MALT have had previous *H. pylori* infection.[134] In fact, *H. pylori* eradication confers a ~74% remission rate of MALT in western populations.[135] The data on the association between *H. pylori* and gastric MALT were very limited. We used the measure of association (OR = 6.3, 95% CI: 2.0–19.9) reported in a study that combined data from two cohort studies conducted in Norway and the USA.[122] This study included 33 cases matched to four controls by cohort, sex, age, and sera collection date.[122]

EPSTEIN-BARR VIRUS (EBV)

EBV was first isolated in 1964 in cells derived from Burkitt lymphoma [136]. Infection often occurs in childhood and presents no symptoms.[137] Virtually all adults are infected with EBV; 90% or more of adults had EBV by 1975 as reported by the IARC.[33] This virus is primarily transmitted orally,[1] but can also be acquired via genital transmission, transfused blood products, stem cell or organ donation.[138] Carcinogenicity is demonstrated by the detection of EBV viral genome within the tumour cells.[139] EBV-encoded RNA *in situ* hybridization (EBER ISH) detection of EBV in tumour cells, considered the gold standard and the most reliable assay to detect EBV in cancer tissues.[1,120] Latent

membrane protein (LMP1) is comparable to EBER for detecting EBV in Hodgkin lymphoma.[120] The high relative risks of the infection-cancer association and the "gold standard" detection of EBV within tumour tissues permitted utilizing the prevalence of EBV in cancer cases to approximate the PARs.

Burkitt lymphoma

The studies identified in our searches were ineligible because they were conducted in Africa or Asia and/or or involved pediatric patient populations. We then ran a more targeted search in PubMed using the terms: "Epstein-Barr virus", "Burkitt lymphoma", and "United States" or "Canada" or "North America" and found one eligible study. Mbulaiteye et al. used EBER ISH to detect EBV infection in 74 cases diagnosed from 1979–2009 using SEER data collected from Los Angeles County, Hawaii, and Iowa.[140] The prevalence/PAR of EBV was 40.0% (95% CI: 22.7–59.4%) among those less than age 50 and 28.6% (95% CI: 11.3–52.2%) for those 50 and older. We calculated Fisher exact 95% CIs were calculated in an online open source tool.[141]

Extranodal natural killer T-cell lymphoma (ENKTL) – nasal type

ENKTL – nasal type is a rare aggressive form of non-Hodgkin lymphoma impacting the palate and nasal fossa, can, although rarely, affect the skin and digestive tract.[142] It is exceptionally rare in Europe and North America accounting for less than 0.1% of non-Hodgkin lymphomas.[143] It has been established that EBV is detected in virtually all cases of ENKT nasal type.[24] Our literature search did not reveal evidence to the contrary. Hence, we attributed all 25 ENKTL – nasal type cases diagnosed in Canada in 2015 to EBV.

Hodgkin lymphoma (HL)

The proportion of EBV-positive HL depends on the geographic region, age, histological type, and immune status of infected individuals.[24] In particular, the proportion of lymphomas that are EBV-related substantially varies by region.[56] The virus is most often associated with classic HL.[1] The latter accounts for 95% of all HL, according to the Canadian Cancer Society.[144] Our literature search identified one relevant meta-analysis on the prevalence of EBV in HL;[145] the pooled prevalence from 12 North American studies was 31.8% (95% CI: 25.3–39.1). However, only four of these studies (**Table S7** and **Fig. S2**) met our inclusion criteria; the rest were excluded (pediatric study populations, data not reported or available upon requesting the prevalence of EBV in HL by sex).

Table S7. Studies reporting on Epstein-Barr virus prevalence in Hodgkin lymphoma tumor tissues in North American populations by sex

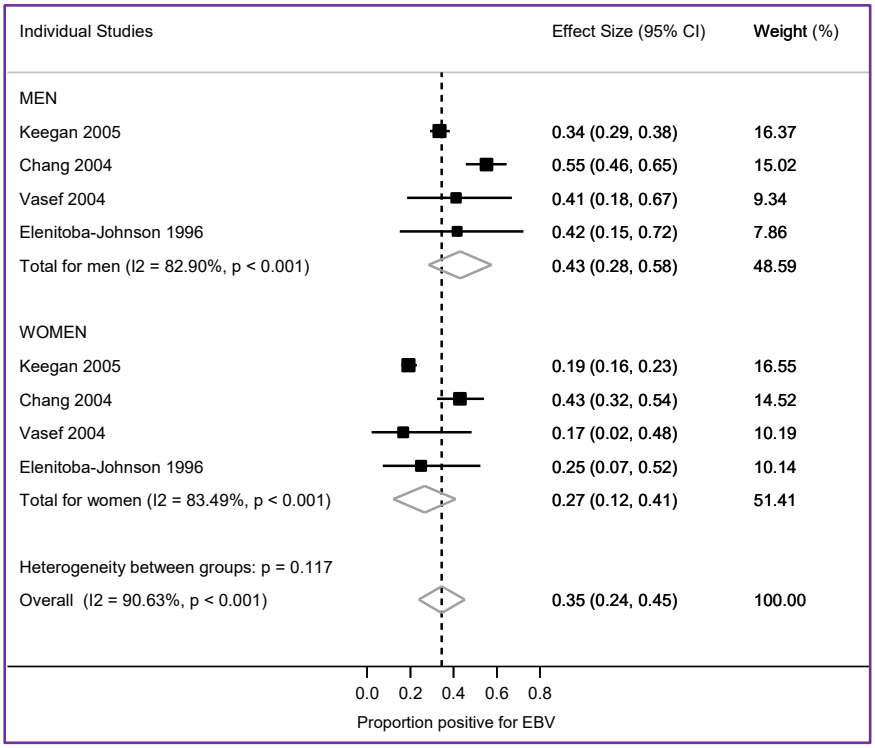
Study[a]	Region	Sources of cases	Period of analysis	Detection method(s)	Men		Women	
					Cases N	Positive %	Cases N	Positive %
Keegan 2005[146]	California, US	Cancer registry	1988–1997	EBER ISH, LMP1	417	33.6	469	19.2
Chang 2004[147]	Massachusetts and Connecticut, US	Population-based case control study	1997–2001	EBER ISH, LMP1	114	55.3	86	43.0
Vasef 2004[148]	Iowa, US	Pathology department	--	EBER ISH, LMP1	17	41.2	12	16.7
Elenitoba-Johnson 1996[149]	Rhode Island, US	Pathology departments	--	LMP1	12	41.7	16	25.0

EBER ISH = Epstein-Barr encoding region *in situ* hybridization; LMP1 = latent membrane protein 1, US = United States

-- Indicates that it was not reported in the original study.

a. Inclusion criteria: tissue specimen tested for EBV, EBER ISH or LMP1 detection, Canadian or American participants aged 15 and older, EBV results available by sex (in text or by request), 10 or more participants.

Fig. S2. Forest plot of EBV prevalence in Hodgkin lymphoma, by sex



Nasopharyngeal cancer (NPC)

NPC, a tumour of the epithelial tissues, is classified into two main types, non-keratinizing squamous cell accounting for 80% of all NPCs and the keratinizing type accounting for the remaining 20%.[150] The pooled prevalence of the two included studies reporting on EBV in NPC was 69.4% (95% CI: 61.9–76.9%) indicating that the majority of NPC are due to EBV infection (**Table S8**).

Table S8. Characteristics of studies reporting on Epstein-Barr virus prevalence in North American nasopharyngeal carcinoma patients

Study[a]	Region	Source of cases	Diagnosis dates	Detection method	Keratinizing %	Tested n	Positive %	95% CI	Weight %
Dogan 2014[b](151)	Pennsylvania, Washington, US	Pathology archives	1981–2012	EBER ISH	14.3	63	60.3	47.2–72.4	38.15
Shi 2002(152)	Ontario, CA	Hospital	1985–1992	EBER ISH	18.8	80	75.0	64.1–84.0	61.85

CA = Canada, EBER ISH = Epstein-Barr encoding region *in situ* hybridization, US = United States

a. Inclusion criteria: tissue specimen tested for EBV, EBER ISH detection, Canadian or American participants aged 15 and older, and 10 or more participants.

b. EBV positivity was reported for two periods of diagnoses: 1956–1977 and 1981–2012, cases from the first diagnoses period were excluded because they occurred 59 to 38 years before 2015 (the year we are applying PARs to).

Human papillomavirus (HPV)

HPV was first reported as carcinogenic in the IARC's 1995 Monograph (vol. 64).(153) A comprehensive review was published in 2007 (vol. 90),(26) and the most recent update was published in 2012 (vol. 100B).(1) Thirteen of the more than 200 identified HPV types are established carcinogens.(1,154) Papillomaviruses contain 16 genera and are part of the *Papillomaviridae* family. The alphapapillomavirus genus contains the papilloma types that infect the mucosa potentially leading to mucosal tumours. The virus, transmitted through skin-to-skin contact, is the most common sexually transmitted infection and is highly prevalent among sexually active individuals.(155) Oral HPV infection is mainly spread through open-mouth kissing and oral sex.(47) About 90% of HPV infections clear spontaneously within two years, but immune clearance of the virus is less likely in people with compromised immune systems.(156,157) The most recent monograph classified HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59 as known carcinogens, also referred to as Group 1 carcinogens.(1) HPV68 is considered 'probably' carcinogenic (Group 2A), and several HPV types (26, 30, 34, 53, 66, 67, 69, 70, 73, 82, 85, 97) as 'possibly' carcinogenic (Group 2B).(1)

Anogenital cancers

Persistent HPV infection is the strongest risk factor for anal, penile, vaginal, and vulvar cancers, with all cervical cancers being caused by HPV infection.(158) Although IARC has reported that 12 HPV types are convincingly linked to cervical cancer, there is 'sufficient' evidence for carcinogenicity in other anogenital sites has only been established for HPV16.

Anal cancer

When pooling six studies (**Table S9**, **Fig. S3**) that met our inclusion criteria, the analysis automatically excluded the study by Meyer et al., 2013 due to inadmissible CIs (exceeding a proportion of 1.0). We thus enabled the "ftt" command, forcing the upper bound to 1.0 in order to include this study in the analysis.

Table S9. Characteristics of studies reporting on high-risk human papillomavirus prevalence in invasive anal cancers in North American populations by sex

Study[a]	Region	Source of cases	Diagnosis dates	Detection methods high-risk types tested[b]	Specimen	Men		Women	
						Tested N	Positive %	Tested N	Positive %
Chung 2016[(159)]	Massachu-setts, US	--	2013–2014	HC-based sequencing 16, 18	FFPE	22	77.3	48	91.7
Alemany 2015[(160)]	Multiple regions, US	Pathology archives	1986–2011	SPF-10 PCR, DEIA, $LiPA_{25}$ 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 58, 66, 68, 70, 97	FFPE	39	79.5	57	94.7
Ouhoummane 2013[(161)]	Quebec, Canada	Hospitals[c]	1990–2005	LA PGMY primers 16, 18, 33, 53, 56, 58	PE	33	75.8	63	98.4
Meyer 2013[(162)]	New York, US	Surgical pathology files[c]	1997–2009	SPF10 PCR-DEIA-$LiPA_{25}$ 16, 18, 31, 33, 35, 39, 45, 51, 56, 58, 59, 66, 73	--	23	100.0	19	100.0
Herfs 2017[(163)]	Arkansas, Boston, US	Pathology archives[c]	2001–2015	PCR-RT 16, 18, 31, 33, 35, 39, 45, 51, 56, 58, 59, 66, 68	FFPE	23	91.3	27	88.9
Steinau 2013[(164)]	Florida, Hawaii, Iowa, Kentucky, Lousiana, Michigan, Los Angeles County, US	Cancer registries, tissue repositor-ies	1995–2005	PCR, LA, INNO-LiPA 16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 54, 56, 58, 59, 66, 68	FFPE	53	84.9	93	90.3

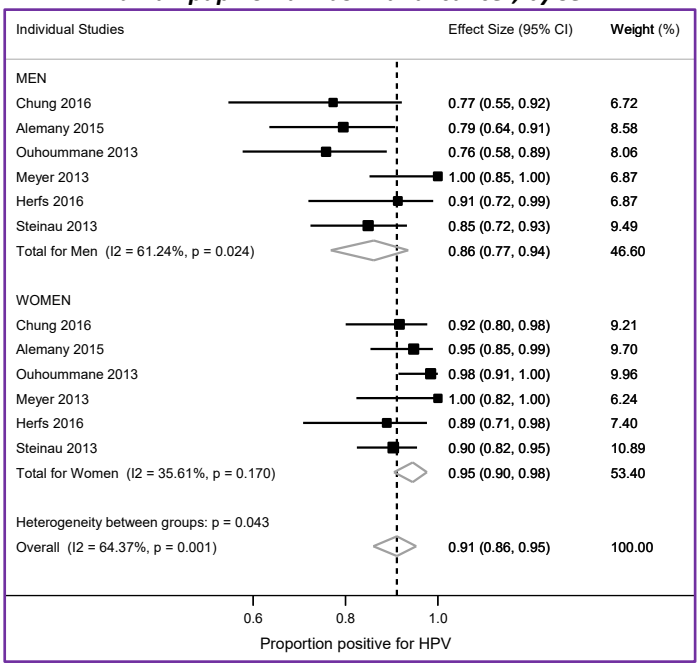
FFPE = Formalin-fixed paraffin-embedded, PCR = polymerase chain reaction, PE = paraffin-embedded, LA = Linear Array, HC = Hybrid Capture, RT = "RealTime", US = United States

a. Inclusion criteria: invasive anal cancers tissue specimens, PCR detection, 10 or more cases, North American study population, published after 1995, data stratified by sex or available upon request.

b. High-risk HPV types included: 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 59, 68, 66, 70, 73 and 97.

c. Included some cases known to be HIV positive.

Fig. S3. Forest plot of the prevalence of high-risk human papillomavirus in anal cancer, by sex



Penile cancer

Pooling the six studies (**Table S10** and **Fig. S4**) that met the inclusion criteria provided a PAR of 39.4. Heterogeneity was very high was an I^2 of 91.7%.

Table S10. Characteristics of studies reporting on high-risk human papillomavirus prevalence in invasive penile cancers in North American populations

Study[a]	Region	Source of cases	Diagnosis dates	Detection methods[b] high-risk types genotyped[c]	Specimen	Tested N	Positive %
Alemany 2016[165]	Hawaii[d], Iowa, US	Pathology archives	1994–2004	SPF-10, DEIA, LiPA$_{25}$ 16, 18, 31, 32, 33, 35, 39, 45, 51, 52, 56, 58, 59	FFPE	16	18.8
McDaniel 2015[166]	Michigan, US	Pathology archives	2005–2013	GP5+/6+, MY09/11 16, 33	FFPE	43	11.6
Hernandez 2014[167]	Kentucky, Louisiana, Michigan, Iowa, Hawaii[d], Los Angeles County, US	Population-based cancer registries, residual tissue repositories	1998–2005	PCR, LA, LiPA$_{25}$ 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68	FFPE	79	59.5
Daling 2005[168]	Western Washington, US	Population-based cancer registry	1979–1998	PCR-MY09/11, L1 16, 18, 31, 33, 45, 53	PE	43	62.8
Rubin 2001[169]	Michigan, US	Pathology archives	Not reported	PCR SPF 10, LiPA$_{25}$ 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70	FFPE	88	39.8
Cupp 1995[170]	Minnesota, US	Pathology archives	1981–1993	PCR L1, PCR-E6, TS 16, 18	FFPE	42	42.9

PCR = polymerase chain reaction, FFPE = formalin-fixed paraffin-embedded, PE = paraffin-embedded LA = linear array, TS = type specific, US = United States

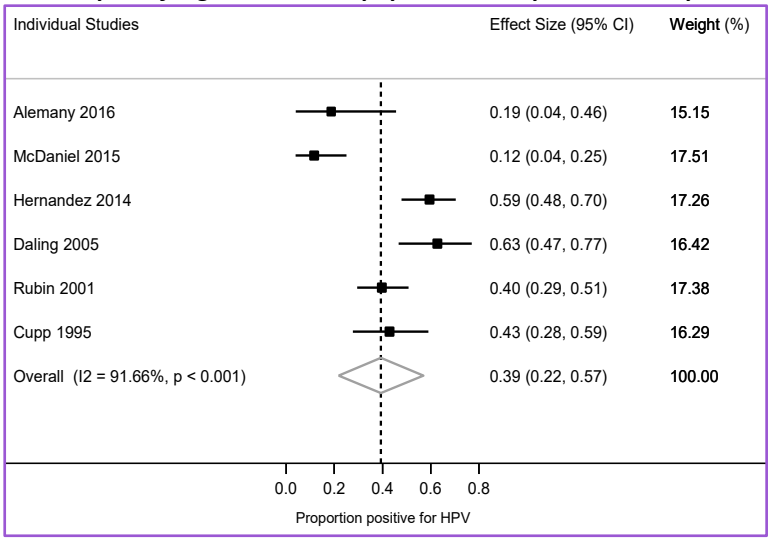
a. Inclusion criteria: invasive penile cancers tissue specimens, PCR detection, 10 or more cases, North American study population, published on or after 1995

b. All HPV testing was performed with polymerase chain reaction (PCR).

c. High-risk HPV types included: 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 59, 68, 66, 70, 73 and 97.

d. 3 cases overlapped.

Fig. S4. Forest plot of high-risk human papillomavirus prevalence in penile cancer



Vaginal cancer

Two studies met our inclusion criteria (**Table S11**). The pooled prevalence of high-risk HPV types in invasive vaginal cancers was 72.2% (95% CI: 62.8 – 81.7). The Sinno 2014 study had 14% of cases with non-SCC cancer types.

Table S11. Characteristics of studies reporting on high-risk human papillomavirus prevalence in invasive vaginal cancers in North American populations and results of pooled analysis

Study[a]	Region	Source of cases	Diagnosis dates	Detection methods high-risk HPV[b] types tested	Specimen	Tested N	Positive %	95% CI	Weight %
Sinno 2014[171]	California, Florida, Hawaii, Kentucky, Louisiana, Iowa, Michigan, USA	Population-based cancer registries, residual tissue repositories	1994–2005	LA, INNO-LiPA 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 67, 68, 70, 73, 82	FFPE	60	75.0	62.1–85.3	74.7
Daling 2002[172]	Washington state, USA	Population-based cancer registry	1981–1998	PCR-L1, MY09/MY11 16, 18/45, 31	PE	25	64.0	42.5–82.0	25.3

LA = linear array, PCR = polymerase chain reaction, PE= paraffin-embedded, FFPE = formalin-fixed paraffin-embedded

a. Inclusion criteria: invasive vaginal cancer tissue specimens, PCR detection, 10 or more cases, North American study population, published on or after 1995.

b. High-risk HPV types included: 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 59, 68, 66, 70, 73 and 97.

Vulvar cancer

There were two studies per age group (**Table S12**). The pooled prevalence of HPV in cases was higher for younger women (76.8%) compared to older women (43.2%) (**Fig. S5**). As HPV is more prevalent among younger vulvar cancer cases, one PAR was calculated for women <50 years of age and one for those older.

Table S12. Characteristics of studies reporting on the prevalence of high-risk human papillomavirus in invasive vulvar cancer cancers in North American populations by age group

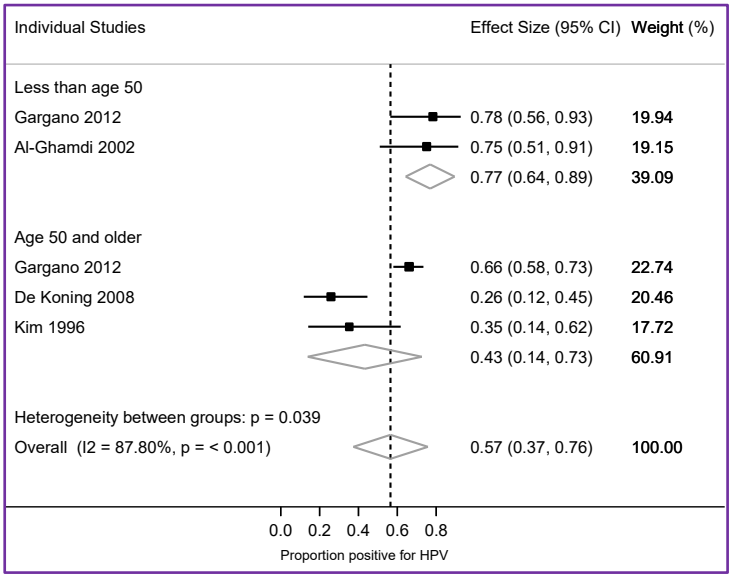
Study[a]	Region	Source of cases	Diagnosis dates	Detection methods high-risk HPV[b] types tested	Specimen	Age <50 years		Age ≥50 years	
						Tested N	Positive %	Tested N	Positive %
Gargano 2012[173]	California, Florida, Hawaii, Iowa Kentucky, Louisiana, Michigan, US	Population-based cancer registries, residual tissue repositories	1995–2005	LA, INNO-LiPA 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 67, 68, 70, 73, 82	FFPE	23	78.3	153	66.0
de Koning 2008[174]	New York, US	Pathology department	1990–2005	SPF10, $LiPA_{25}$ 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70	PE	--		31	25.8
Al-Ghamdi 2002[175]	British Columbia, Yukon, CA	Population-based cancer registry	1970–1998	PCR-MY09/11, PCR-GP5/6 16, 18	FFPE	20	75.0	--	
Kim 1996[176]	Maryland, Florida, US		1989–1994	PCR-MY09/11, PCR L1, TS, Sequencing 16, 18	fresh	--		17	35.3

CA = Canada, LA = linear array, PCR = polymerase chain reaction, PE= paraffin-embedded, FFPE = formalin-fixed paraffin-embedded, US = United States

a. Inclusion criteria: invasive vulvar cancers tissue specimens, PCR detection, 10 or more cases, North American study population, published on or after 1995, data stratified by age or available upon request.

b. High-risk HPV types included: 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 59, 68, 66, 70, 73 and 97.

Fig. S5. Forest plot for high-risk human papillomavirus prevalence in vulvar cancer by age group



Head and neck cancers (HNCs)

Whereas we used the prevalence of high-risk HPV types detected via PCR techniques to approximate PARs in anogenital cancer sites, attributing HNCs to HPV requires detecting the oncoproteins E6 and E7, as it is recognized as the gold standard.[177,178] The oncoproteins E6/E7 are produced by high-risk HPV and must be present for viral replication to occur. The prevalence in cases as detected by oncoproteins approximates the PAR because the RRs between HPV and HNCs have consistently reported very strong relative risks.[47-49] We only considered the prevalence of HPV16 as the association between HNCs and HPV is most established for this type. We assumed that the cases were invasive and primary tumors unless otherwise specified. There were 16 studies for the oropharynx, nine for the oral cavity, and five for the larynx (**Table S13**). The PAR was highest for the oropharynx at 60% (**Fig. S6**), followed by the larynx with 13% (**Fig. S8**), then the oral cavity with 8% attributable to HPV16 (**Fig. S7**).

Table S13. Characteristics of studies reporting on HPV type 16 prevalence detected via E6 and/or E7 in head and neck cancers in North American populations

Study[a]	Region[b]	Diagnosis dates	Detection method(s)	Specimen	Anatomical site[c]					
					Oropharynx		Oral cavity		Larynx	
					Tested N	Positive %	Tested N	Positive %	Tested N	Positive %
Biron 2016[179]	Alberta, CA	2015	ddPCR E6 or 7	Fresh	29	72.4	16	0.0	--	
Hooper 2015[180]	Oregon, US	--	PCR-E6, E7	FF	44	68.2	24	8.3	19	0.0
Isayeva 2014[181]	Alabama, US	2005–2012	RT-PCR E6 and 7	PE	102	48.0	--		--	
Nichols 2013[182]	Ontario, CA	2003–2009	PCR-E6, PCR-E7	FFPE	95	47.4	--		--	
Lingen 2013[d][183]	California, Illinois, Ohio, USA Ontario, CA	2005–2011	qRT-PCR E6 or 7	FFPE	--		409	3.7	--	
Walline 2013[184]	Michigan, US	2001–2011	PCR-E6	FFPE	208	78.9	104	4.8	--	
Jordan 2012[185]	California, Illinois, Ohio, USA Ontario, CA	2000-2009	qPCR E6	FFPE	235	62.1	--		--	
Stephen 2012[186]	Michigan, US	1999–2007	qRT-PCR E6	FFPE	--		--		77	27.3
Chaturvedi 2011[187]	Hawaii, Iowa, Los Angeles, California, US	1984–2004	qRT-PCR E6	FFPE	216	35.2	--		--	
Schlecht 2011[188]	New York, US	--	TS-PCR E6 or E7	FF, PE	23	52.2	29	27.6	27	18.5
Agoston 2010[189]	Massachusetts, US	--	PCR L1 /E7/DNA-DNA ISH	FFPE	126	58.7	--		--	
Jo 2009[190]	California, US	2000–2003	PCR-E7	FF, FFPE	14	92.9	--		--	
Settle 2009[191]	Maryland, US	1995–2006	PCR-E6	PE	119	49.6	28	10.7	55	7.3
Tezal 2009[192]	New York, US	1999–2005	TS-PCR E6	PE	30	70.0	--		--	
Cohen 2008[193]	Pennsylvania, US	1996–2001	TS-PCR E7	PE	35	68.6	--		--	
Liang 2008[194]	Minnesota, US	2004–2006	TS-PCR E6	FF	--		51	2.0	--	
Worden 2008[195]	Michigan, US		RT-PCR E6		42	64.3	--		--	
Zhao 2005[196]	Maryland, US	1984–2002	RT-PCR E6/E7	Frozen	26	57.7	38	15.8	16	18.8
Ha 2002[197]	Maryland, US	1982–2000	E6 and 7 via TaqMan	23 FF 11 PE	--		34	20.6	--	
Strome 2002[198]	Minnesota, US	1987–1995	TS-PCR E6	PE	52	40.4	--	--	--	--

CA = Canada, ddPCR = droplet digital PCR, FFPE = formalin-fixed paraffin embedded, PE = paraffin embedded, FF = fresh-frozen, TS = type-specific, qRT-PCR = real-time quantitative reverse transcription PCR, US = United States

-- Indicates the cancer was not included in the original study or that it overlapped with another included study.

a. Inclusion criteria: site specific results (e.g. base of tongue versus oral tongue), detection in cancer tissue, invasive and untreated cancer, detection with E6 and/or E7 for HPV16, North American study population and published in 2000 or later (evidence that indicates that PCR technique had not been refined in the 1990s and the HPV-associated HNCs increased over time.[89] Did not test specimens for E6/7 based on previous HPV results (for example, positive for HPV via PCR then sent to E6/7).

b. Only cases from Chaturvedi et al.'s 2011 study originated from population-based cancer registries, the remaining studies cases came from clinics, hospitals, and pathology departments.

c. Tested positive for E6 and/or E7.

d. Lingen 2013 included some in situ cases.

Fig. S6. Forest plot of human papillomavirus type 16 E6/E7 prevalence in oropharyngeal cancer

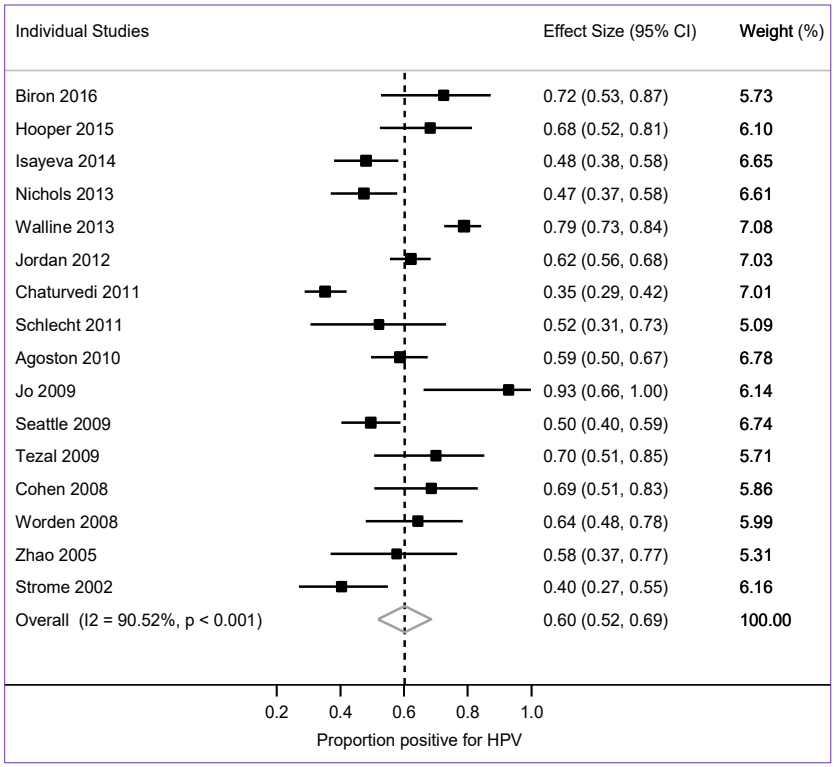


Fig. S7. Forest plot of human papillomavirus type 16 E6/E7 prevalence in oral cavity cancer

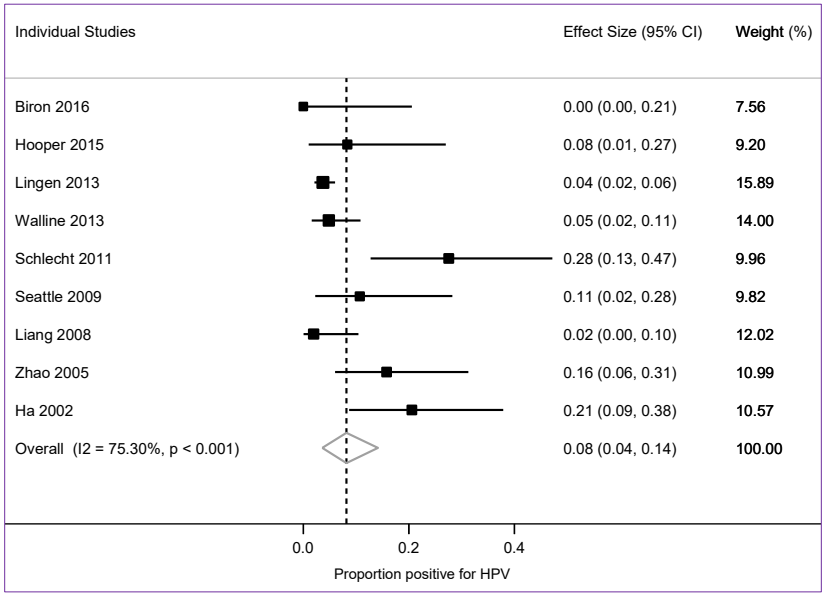
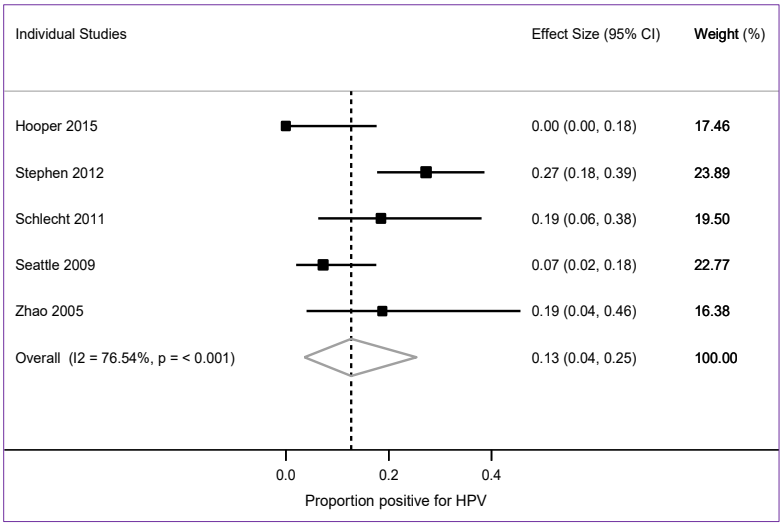


Fig. S8. Forest plot of human papillomavirus type 16 E6/E7 prevalence in laryngeal cancer



Human herpesvirus, type 8 (HHV-8)

HHV-8, frequently referred to as Kaposi's sarcoma-associated herpes virus, is a large double-stranded DNA gamma herpes virus.[199] HHV-8 is a necessary, but not sufficient cause of Kaposi sarcoma. The virus is also implicated in primary effusion lymphoma. There four types of KS, epidemic (AIDS related), classic (Mediterranean), endemic (African), and latrogenic (transplant related). In Canada and the United States, epidemic/HIV related is the most common of the Kaposi sarcoma type followed by latrogenic/transplant related. Previously known as body cavity lymphoma, primary effusion lymphoma is an aggressive and rare subtype of diffuse large B-cell non-Hodgkin lymphoma. All patients with primary effusion lymphoma have HHV-8 and many also have EBV. All Kaposi sarcomas and primary effusion lymphomas diagnosed in Canada were attributed to HHV-8.

Human T-cell lymphotropic virus, type 1 (HTLV-1)

HTLV was the first human retrovirus discovered in 1977 in the United States; and, independently, the virus was also found in Japan.[121] HTLV has four types, but only HTLV type 1 is linked to cancer.[121] In Canada, screening of the blood supply commenced in 1990 for antibody to HTLV-1.[200] HTLV is not nationally reportable and there are no general population prevalence estimates. However, among 518,957 Canadians donating blood for the first from 2005–2010, 46 tested positive for HTLV-1[200] (personal communication, Dr. Sheila O'Brien). Although blood donors are a selected population, this

represents a prevalence of less than 0.001%. In virtually all cases of adult T-cell leukemia/lymphoma, HTLV-1 is present.[201] Therefore, we attributable all 30 cases of adult T-cell leukemia/lymphoma diagnosed among Canadians in 2015 to HTLV-1.

We are grateful to those who provided us with additional data:

Dr. Sam Mbulaiteye (Division of Cancer Epidemiology & Genetics, National Cancer Institute)
Dr. Julia Gargano (HPV Team/Viral Vaccine-Preventable Diseases Branch/Division of Viral Diseases/NCIRD/CDC)
Dr. Edyta C. Pirog (Weill Cornell Medicine, Cornell University)
Dr. Margaret Madeleine (Program in Epidemiology, Fred Hutchinson Cancer Research Center)
Dr. Janet Daling (Program in Epidemiology, Fred Hutchinson Cancer Research Center)
Dr. Michael Herfs (Laboratory of Experimental Pathology, GIGA-Cancer, University of Liège)
Dr. Jon H. Chung (Clinical development, Foundation Medicine)
Dr. Ping Yan (Public Health Agency of Canada, Government of Canada)
Dr. Max Trubnikov (Department of Indigenous Services, Government of Canada)
Dr. Chris Archibald (Public Health Agency of Canada, Government of Canada)
Dr. Sheila O'Brien (Canadian Blood Services)

Manuscript #2: Cancers attributable to infections in the United States in 2017

This manuscript provides estimates on the impact of seven infections on cancer incidence in the US in 2017. HIV is also considered in the estimates relating to EBV-related lymphomas. It includes an analysis for both adults and children (for EBV only). While the first manuscript included only established carcinogens and cancers where the evidence for the role of the infection was deemed 'sufficient' by the IARC (with the exception of laryngeal cancer), this manuscript includes PAF estimates for seven additional infection-cancer pairs.

This manuscript is not yet formatted for a specific journal.

Cancers attributable to infections in the United States in 2017

Karena D. Volesky[1,2]
Samantha Morais[1]
Stephen D. Walter[3]
Thomas R. O'Brien[4]
Allan Hildesheim[4]
Eric A. Engels[4]
Mariam El-Zein[1]
Eduardo L. Franco[1,2]

1. Division of Cancer Epidemiology, McGill University, Montreal, Canada
2. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada
3. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
4. Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, United States

Data availability: The relevant data are available within the manuscript and its supplement. However, cancer case counts below 16 were aggregated with other cancers or by sex/age group because the Surveillance, Epidemiology, and End Results Program (SEER) Research Data Use Agreement advises against the publication of case counts below 16.

Financial support: This work is supported by a foundation grant (FDN-143347) to ELF and a doctoral award (support for KDV) jointly provided by Fonds de recherche du Québec – Santé (FRQS) and Fondation des étoiles.

Acknowledgements: We appreciate Genevieve Gore's (Liaison Librarian at McGill University) assistance in the development of the search strategy. Thank you to Dr. Yibing Ruan (Alberta Health Services, Canada) who provided valuable mentorship to KDV on earlier work that estimated the infection-attributable cancer burden in Canada.

ABSTRACT

Infections are important, largely modifiable, causes of cancer. To estimate the impact of infections on cancer incidence in the United States, we calculated population attributable fractions (PAFs) for 31 infection-cancer pairs. Data from 125 studies were meta-analyzed to obtain the magnitude of an infection-cancer association or prevalence of the infection within cancer cells. The National Health and Nutrition Examination Survey data were used to obtain population prevalence estimates for infections with hepatitis B and C viruses, and *Helicobacter pylori*. Of the 1,662,102 cancers diagnosed among individuals aged ≥20 years in the United States in 2017, 71,469 (4.3%; 95% confidence interval [CI]: 3.2–6.2%) were attributable to seven infections. The cancers with the highest number of infection-attributable cancers were cervical (human papillomavirus [HPV], n = 12,829), oropharyngeal (HPV, n = 12,599), and non-cardia gastric (*H. pylori*, n = 11,766). The burden of infection-attributable cancers as a proportion of total cancer incidence was highest for females and males aged 20–49 years (6.5% and 7.4%, respectively), followed by males and females aged ≥50 years (4.3% and 3.6%, respectively). HPV accounted for more than half (53.8%) of infection-attributable cancers, followed by *H. pylori* (16.6%), hepatitis C virus (12.8%), Epstein-Barr virus (11.1%), hepatitis B virus (3.3%), human herpesvirus type 8 (1.5%), finally human T-cell lymphotropic virus type 1 (0.9%). Among those aged 0–19 years old, 2.2% (95% CI: 1.3–3.0%) of cancers diagnosed in 2017 were attributable to Epstein-Barr virus.

1. INTRODUCTION

Public awareness of the role of infections in cancer etiology is low.[202] Yet, the strongest causal relationships in cancer etiology are those involving certain infections.[42,77] In fact, certain infections are the sole causes of cancer to be deemed necessary.[1,4,5,78] Specifically, human papillomavirus (HPV) and human herpesvirus type 8 (HHV-8) are necessary causes of cervical cancer and Kaposi sarcoma, respectively.[1,4,5,78] Importantly, infections causing the most cancers globally are preventable (vaccination for HPV and hepatitis B virus [HBV]) or treatable (direct-acting antivirals for hepatitis C virus [HCV] and antibiotic therapy for *Helicobacter pylori* [*H. pylori*]).[56,57] While efforts to reduce the prevalence of these infections in the United States (US) are ongoing,[203] there remains considerable untapped potential for the prevention and treatment of carcinogenic infections.

Islami and colleagues estimated that 3.3% of cancers diagnosed among those aged ≥30 years were attributable to infections in 2014 in the US.[17] However, this study did not include all carcinogenic infections and associated cancers nor included cancers diagnosed among those less than age 30, thereby providing an incomplete portrait of the infection-associated cancer burden in the US. Furthermore, in the more than 10 years since the International Agency for Research on Cancer (IARC) updated its assessment of infections[2] evidence has accumulated on the role of carcinogenic infections in additional cancers.[204-206] Hence, the burden of infection-attributable cancer in the US is likely greater than previously estimated.[17]

The considerable potential for the prevention and treatment of carcinogenic infections, coupled with the lack of comprehensive estimates of the impact of infections on cancer incidence in the US, makes quantifying the infection-attributable cancer burden a priority. The goal of this study was to estimate the percentage and number of cancers attributable to infections in the US.

2. METHODS

2.1. Selection of infections and cancers

The IARC classifies infections as carcinogenic (group 1), probably (group 2A) or possibly carcinogenic (group 2B), or not classifiable (group 3).[24] For the main analysis, we included group 1 infections and associated cancers with 'sufficient' evidence according to the IARC, with three exceptions. First, we excluded parasitic infections (i.e., *Opisthorchis viverrini*, *Clonorchis sinensis*, and *Schistosoma haematobium*) because they do not occur in endemic form in the US;[207-209] however, we recognize the possibility that immigrants from countries where these parasites are endemic, remain

at risk of parasitic-related cancers (e.g., bladder cancer and cholangiocarcinoma). Second, we did not consider the role of human immunodeficiency virus (HIV) in cancer of the conjunctiva (cancer with 'sufficient' evidence) because this cancer is very rare in the US and we lacked the data required to include this association.[210] Third, since some non-Hodgkin lymphoma (NHL) subtypes, more clearly demonstrate a relationship with HCV,[1,123,211,212] NHLs were analyzed by subtype and not as a single entity.

The main analysis was extended in several ways. Cancer associations with 'limited' evidence – each HBV and HCV and cholangiocarcinoma (i.e., cancer of the bile ducts) were included because several meta-analyses have reported increased risks associated with HBV and HCV.[213-216] Due to differing magnitudes of association between each HBV and HCV and cholangiocarcinoma arising in the intrahepatic versus extrahepatic bile ducts, these two subsites were analyzed separately. We included cancer of the larynx (cancer with 'limited' evidence) because there is broad support for the etiologic role of HPV in a small fraction of laryngeal cancers.[83,84] We included two cancers where Epstein-Barr virus (EBV) is believed to play an etiologic role – diffuse large B-cell lymphoma (DLBCL) and gastric cancer.[206,217,218] These cancers were selected because their inclusion has the potential to increase the burden of infection-attributable cancers among adults, DLBCL is the most commonly diagnosed histologic type of NHL[219] and gastric cancer is in the 15 most commonly diagnosed cancers in the US.[220] We included *H. pylori* and esophageal cancer because several meta-analyses have reported an inverse association.[221-226] Accounting for the protective effect of *H. pylori* will provide a more accurate estimate of the impact *H. pylori* has on cancer incidence. Finally, the role of EBV in cancers diagnosed among children and adolescents (aged 0–19 years, herein referred to as children) was included because EBV is an established cause of a proportion of lymphomas arising in children.[1] Other infection-related cancers are extremely rare among children and therefore were not considered.

Through immunosuppression, HIV amplifies the carcinogenic effects of infections such as EBV and HPV.[1] Analytically, we considered HIV as a modifier rather than a direct cause and did not attribute cancers directly to HIV. Where data permitted, we calculated separate estimates for people living with HIV (PLWH). Since we do not have recent cancer incidence data for PLWH, we did not summarize the impact of infections on cancer incidence in this group; we refer readers interested in such data to a paper published by de Martel and colleagues in 2015.[227]

2.2 Population attributable fractions (PAFs)

Formulas for calculating PAFs for binary exposures

Formula 1

$$PAF = \frac{Pe(RR-1)}{1+Pe(RR-1)}$$

Formula 2

$$PAF = Pc\frac{(RR-1)}{(RR)}$$

Formula 3

$$PAF = Pc$$

PAFs (estimated via three alternative equations) represent the proportion of cancer incidence associated with the exposure. Formula 1 requires prevalence of the infection in the general population (Pe) and its relative risk (RR) associated with the cancer;[69] formula 2 can estimate PAFs using prevalence in cases (Pc) instead of Pe;[72] and, formula 3 can be used when the attributable fraction in the exposed group approaches 1.0 (i.e., RRs or ORs are high), such that the prevalence in cases approximates the PAF and/or when mechanistic evidence exists for the role of the infection in cancer thereby allowing the PAF to be approximated by the prevalence in cancer cases.[1,47,56] We estimated the PAFs for HBV, HCV and *H. pylori* via formula 1, and the remaining infections via formula 3.

2.3. Data acquisition

To obtain data for the PAF calculations, we searched IARC monographs,[1,25,26,85,126,228] the Catalan Institute of Oncology HPV Information Centre report for the US,[229] other PAF analyses,[17,45,46,54,56,57,227] contacted experts, and performed a literature search. The purpose of the literature search was to identify knowledge syntheses (systematic reviews, integrated reviews, meta-analyses, etc.) from which we could identify individual studies for the PAF inputs. The search, detailed in **Supplementary Table 1**, was conducted in MEDLINE (1946–) on September 15, 2021. It included MeSH terms and keywords related to infections and cancers included in this study, and knowledge syntheses, and was limited to records published in English. For each included record, KDV or SM performed a forward citation search (i.e., identifies studies citing the specific article, which were then reviewed for eligibility), and the references of included individual studies were reviewed for potential records. For the burden of EBV-attributable cancers in children, studies were identified from an ongoing systematic review and meta-analysis by our team, for which a search was performed in Embase and MEDLINE (PROSPERO protocol: CRD42021269730).

2.4. Data selection

Table 1 lists the infections and cancers considered, as well as overall and specific inclusion criteria applied for the selection of PAF inputs in terms of prevalence and risk estimates. Across all infections, the cancer cases had to be primary (recurrent cases were excluded), invasive, and not yet treated before the specimen (either serum or cancer tissue) was taken and tested for the carcinogenic

infection. If a study did not clearly report whether the specimen was collected prior to treatment, we assumed it was. We opted to select studies conducted in North America; however, obtaining relevant data for several associations (*H. pylori* and non-cardia gastric cancer [NCGC], EBV and DLBCL, EBV-associated Hodgkin lymphoma arising in PLWH, and EBV in Burkitt lymphoma diagnosed in children) necessitated the inclusion of studies conducted in other Western countries. When the infection is a necessary cause, part of the diagnostic criteria, or widely accepted as the universal cause of a specific cancer, 100% of cases were attributed to the associated infection.

To reflect a biologically plausible latency period between exposure measurement and cancer diagnoses, population prevalence data to input into formula 1 had to be collected prior to the year for which the most cancer incidence data were available (2017).[70] In contrast, when using PAF formula 3, where the PAF is approximated by the prevalence of the infection in cancer tissue (biopsy or surgical specimen), we aimed to select studies that enrolled patients closer to when the cancers were diagnosed.

KDV or SM extracted data, then verified each other's extractions. Authors of studies that met the inclusion criteria but did not report the specific numerator and denominator of interest were contacted for these data.

2.5. Prevalence of HBV, HCV, and H. pylori

The National Health and Nutrition Examination Survey (NHANES) was the source of the US population infection prevalence estimates (HBV, HCV and *H. pylori*) because, when weighted, it is representative of the resident civilian non-institutionalized US population.[230] NHANES tested participants' (aged ≥6) sera for hepatitis B surface antigen (HBsAg) – the marker of active (current) infection and anti-HCV antibodies (marker of past or current infection). Samples testing anti-HCV positive or indeterminate are then tested for HCV RNA – the gold standard marker for active infection.[126] Since only <0.5% of the US population tests positive for HBsAg and <2.0% for HCV RNA, six cross-sectional NHANES cycles (data collected: 1999–2000, 2001–2002, 2003–2004, 2005–2006, 2007–2008, 2009–2010) assessing HBV and HCV prevalence with the same methods, were combined for greater precision.[94,95,103] After combining, the data are representative of the mid-point (2004–2005) of the combined years.[230] *H. pylori* serological status was assessed via enzyme-linked immunosorbent assay (ELISA) in the 1999–2000 NHANES cycle.

Since the HBV, HCV, and *H. pylori* NHANES data were not missing completely at random (i.e., missing values depend on the outcome and covariates, observed or not) multiple imputations with chained equations (25 imputed databases) were performed to minimize possible bias and maximize the available data. The imputation model included variables known to be associated with both the infection and missingness, as applicable (i.e., for all three infections: sex, age [missing age at medical examination was imputed using age at interview when available – last observation carried forward method], education, race, and primary sampling units and strata; HBV infection also included country of birth, intravenous drug use, men who have sex with men, and number of lifetime sexual partners; HCV infection also included injection drug use, receiving a blood transfusion before 1992, HIV diagnosis and anti-HCV antibody result; *H. pylori* infection also included time living in the US, number of people living in the household and family income).[231] We then estimated the prevalence of HBV, HCV and *H. pylori* infection; analyses included the sampling weights provided by NHANES to account for unequal probabilities of selection resulting from the sample design. The recommended variance estimation of Taylor series linearization for variance estimation was used to calculate 95% confidence intervals (CIs) for the prevalence estimates.[232]

The weighted and imputed data were used in the analyses, but for comparison purposes, the imputed data are displayed side-by-side with complete infection prevalence data in **Supplementary Tables 2, 3, and 8**.

2.6. Data analysis

Meta-analytic techniques were used to summarize the measure of association between a given infection and its cancer, and the prevalence of an infection in a given cancer. A fixed effect model was adopted if the index of consistency (I^2) was <25%, and a test for heterogeneity (Cochran Q test) was not statistically significant ($p>0.10$). Data on the number of individuals testing positive and the number with valid testing results (indeterminate results were excluded from the numerator and denominator) were used to calculate pooled prevalence estimates. Pooled prevalence estimates and exact 95% CIs (Clopper-Pearson) were calculated via random effects with the DerSimonian and Laird method, where the Freeman-Tukey double arcsine transformation was used to stabilize variance (enabled only when required).[110,233] When not provided by study authors, OpenEpi[234] was used to calculate measures of association or CIs for proportions. Analyses of NHANES data and meta-analyses were conducted in Stata/SE 17 (StataCorp, College Station, TX, US). PAF calculations and corresponding CIs were

performed in R.[112] The 95% CIs for PAFs computed via formula 1 were calculated with an equation that incorporates uncertainty in both the prevalence and measure of association estimates.[71]

For two infections that can cause the same cancers (EBV and HCV in Burkitt lymphoma; EBV and HCV in DLBCL, and *H. pylori* and EBV in gastric cancer), we assumed that the two infections do not interact and were independent causes of their associated cancers, and therefore we summed their attributable cases. However, for HBV and HCV in hepatocellular carcinoma (HCC) and intra and extrahepatic bile duct cancer, the PAFs for HBV and HCV in HCC were combined with this equation: $1 – (1 – \text{HBV PAF}) * (1 – \text{HCV PAF})$,[72] on the basis that individuals chronically infected with either virus are less likely to proceed to chronic infection if infected with the other virus. To report the impact of HBV and HCV on cancer incidence separately, after combining PAFs with the equation, the number of cases attributable was partitioned by multiplying the proportion of those cases that would be either HBV or HCV if the individual PAFs had simply been summed. To account for *H. pylori's* protective effect in esophageal adenocarcinoma, cases attributable to esophageal adenocarcinoma were subtracted from the total cases attributable to *H. pylori*.

2.7. Cancer incidence

To obtain the number of attributable cases, we multiplied the PAF (proportion), by cancer incidence counts. Cancer incidence data covering 100% of the US population, including the 50 states, the District of Columbia and Puerto Rico (due to Hurricane Maria, Puerto Rico's incidence counts are restricted to the first six months of 2017) were obtained through SEER*Stat software for the most recent year available at the time of analysis, 2017. Specifically, the National Program of Cancer Registries (NPCR) and Surveillance, Epidemiology, and End Results (SEER) Incidence – U.S. Cancer Statistic Public Use Database with Puerto Rico, 2019 submission (2005–2017) was used to obtain the incidence of malignant cancers (*in situ* cases and non-melanoma skin cancers were excluded).[235] Cancer was categorized according to ICD-O-3; we used the following coding classifications: ICD-O-3/World Health Organization (WHO) 2008 for primary sites, lymphoma subsite recode/WHO 2008 for lymphomas, and the International Classification for Childhood Cancer (ICCC) site recode ICD-O-3/WHO 2008 for children. Since case counts below 16 are suppressed by SEER*Stat, we excluded the sex/age-groups with suppressed data for primary effusion lymphoma and nasopharyngeal carcinoma (aged 0–19 only), and for all other cancers we inferred the counts based on totals and then aggregated that

data by sex-age group. PAFs were applied to cancer incidence and the results are shown for males and females ages: 0–19, 20–49, and ≥50 years.

For the main analysis, the NCGC incidence counts were adjusted by reassigning a proportion of 'overlapping lesion' and 'not otherwise specified' (NOS) GC to NCGC. This proportion was determined by calculating the distribution of cardia (C16.0) versus NCGC (C16.1–16.6) by sex and 5-year age groups and multiplying the proportion that was NCGC by the counts of overlapping lesion and NOS, then adding those counts to the existing NCGC counts. We applied the PAFs to unadjusted NCGC (C16.1–16.6) incidence as a sensitivity analysis. For adults, we reclassified B-cell NOS lymphomas based on distribution of B-cell lymphomas of known histology by sex and 5-year age groups, then applied PAFs for EBV to Burkitt lymphoma, EBV and HCV to DLBCL, and HCV to other non-Hodgkin lymphomas.

While we did not attribute any cancers to HIV, the available data permitted calculation of separate PAFs for PLWH and HIV-negative populations for Hodgkin lymphoma and DLBCL. To adjust the number of Hodgkin lymphoma and DLBCL cases attributed to EBV, separate PAFs were applied to cancer incidence partitioned by HIV status. To partition cancer incidence by HIV status, we applied the proportions of Hodgkin lymphoma and DLBCL occurring in PLWH in the US available in the literature.[236,237] The approach for accounting for varying PAFs for PLWH, i.e., EBV and Burkitt lymphoma, Hodgkin lymphoma, and DLBCL, and HPV in anal squamous cell carcinoma (SCC) can be found in the supplement.

Additional background and methodological description for other infection-cancer pairs (HCV and NHLs, EBV and gastric carcinoma, *H. pylori* and its three associated cancers) can be found in the supplement.

3. RESULTS

An overall summary of the PAF inputs for infections where PAFs were estimated using the prevalence of the infection in the population (HBV, HCV, and *H. pylori*) is provided in **Table 2** and for the remaining infections in **Table 3**. The characteristics of individual studies can be found in **Supplementary Tables 4, 6, 7** and **9–21**. The forest plots displaying the pooled measure of associations and prevalence of the infection in cancer can be found in **Supplementary Figs. 1–17**.

Table 4 presents the number of cancers diagnosed and percent attributable to each infection for 22 infection-cancer pairs. Among a total of 1,666,102 cancers diagnosed among adults in the US in

2017, we estimated that the seven infections that we examined were responsible for 4.3% (CI: 3.2–6.2%), translating to 71,469 infection-attributable cancers. Of all cancers diagnosed in adults in 2017 (n = 1,666,102), 2.3% were attributable to HPV (n = 38,468), 0.7% to *H. pylori* (11,881), 0.6% to HCV (n = 9116), and 0.5% to EBV (n = 7942) (Table 4). Among the 27 infection-related cancers listed in Table 4, 33.5% of cases (71,469/213,079) were attributable to infections. HPV's role in anogenital cancers is pronounced for females aged 20–49 years old, who had the highest proportion of cancers due to infections (PAF = 7.4%), largely due to cervical cancer comprising 5.1% of all cancers diagnosed in this group (Table 4). An additional 324 cancers, representing 2.2% (1.3–3.0%) of all cancers diagnosed among individuals aged 0–19, were attributable to EBV; where 77.2% of the cancers attributable to EBV were Hodgkin lymphomas (**Table 5**). While we did not consider cancers attributable to infections other than EBV for children aged 0–19, we note that there were 19 cases of cervical cancer and fewer than 16 cases of each Kaposi sarcoma and adult T-cell leukemia (and hence the count is suppressed) diagnosed in this age group in 2017.

Visualizations of the distribution of infection-attributable cancers among adults by infection, overall and by sex and age group, are provided in **Fig. 1**. While Fig. 1. shows that *H. pylori* occupies an important role across sex and age groups (8.0–21.5% of infection-attributable cancers), the impact of HCV, EBV, and HPV on cancer incidence greatly varied by sex and age. For example, HCV was responsible for only 0.9% of infection-attributable cancers among females aged 20–49 but it accounted for 23.0% of infection-attributable cancers among males ≥50 years old. In individuals aged 20–49 years, EBV was the cause of one-third (31.2%) of infection-attributable cancers in males versus 7.2% in females. HPV in head and neck cancers (HNCs) was responsible from 3.3–34.0% of infection-attributable cancers, while in anogenital sites it ranged from 6.4% in males aged ≥50 to 79.5% in females aged 20–49.

Compared to the weighted NHANES data, prevalence estimates that were both weighted and imputed were higher than the weighted estimates in absolute terms by <0.01% to 0.08% for HCV, 0.01% to 1.5% for HCV, whereas for *H. pylori* they ranged from being 1.11% lower to 2.17% higher, depending on the sex and age group. The combined PAFs for HBV and HCV in HCC greatly varied by sex and age: from 3.4% for females aged 20–39 to 63.7% for males aged 60–64 years. The PAFs for *H. pylori* in NCGC steadily increased with age; from 63.3% for males in their twenties to 82.0% for males aged 75–79, and from 52.2% for females in their twenties to 85.4% for females aged ≥85. The sex and

age group PAFs for HBV and HCV with HCC, and *H. pylori* with NCGC are in **Supplementary Tables 5** and **7**.

We performed several additional analyses (data not shown). We utilized NHL subtype specific ORs for HCV in the main analysis but applied the overall NHL OR of 1.81 (CI: 1.39–2.37) and found that 185 fewer NHLs were attributed to HCV. When the PAFs for *H. pylori* and NCGC were applied to NCGC (ICD-O-3 code: C16.1-16.6) incidence not including reclassified overlapping lesion and NOS gastric sites (which had led to 4,400 more NCGC cases), the number of cases of NCGC attributable to *H. pylori* decreased to 7496 from 11,766. If we attributed 100% of anal SCCs to HPV (as done in a recent global analysis),[57] an additional 365 cases would be attributable to HPV.

4. DISCUSSION

Here, we have reported that ~71,500 (4.3%) of cancers diagnosed among those aged 20 or older in the US in 2017 were attributable to infections; and that ~320 (2.2%) of cancers diagnosed among those aged 0–19 years old were attributable to EBV. Islami and colleagues estimated that 3.3% of cancers diagnosed in the US in 2014 were due to infections.[17] Our estimate is higher due to differing methods. Specifically, we included EBV-associated cancers, intrahepatic and extrahepatic bile duct cancers (HBV and HCV), gastric MALT and DLBCL and esophageal adenocarcinoma (*H. pylori*), and adult T-cell leukemia/lymphoma (HTLV-1); excluding these infections/cancers, our overall PAF estimate would be 3.9%, closer to the estimate reported by Islami and colleagues. Our use of multiple imputation led to higher HBV, HCV, and *H. pylori* prevalence estimates in NHANES data, we also utilized a higher OR for the association between *H. pylori* and NCGC (12.7 versus 5.9 in the analysis by Islami and colleagues). On the other hand, our 4.3% estimate is lower than that reported in a global analysis, which found that 4.8% of cancers diagnosed in the US in 2018 were attributable to infections.[57] Since our analysis included several more infection-cancer pairs than the global analysis, we believe the difference is due to our inclusion of *H. pylori* and esophageal adenocarcinoma and that the PAFs in the global analysis combined infection prevalence for regions comprised of several countries, some of which may have higher infection prevalence than the US.

The demonstrated importance of HPV relative to other infections is consistent with PAF analyses conducted in higher-income countries, such as Australia, Canada, France, and the United Kingdom.[45,58,62,67] Globally, *H. pylori* is the most important infectious cause of cancer,[57] and in Western countries it is the second most important infection after HPV. Accounting for the protective

effect of *H. pylori* in esophageal adenocarcinoma, modestly reduced the overall impact of *H. pylori* on cancer incidence.

4.1. Hepatitis B and C viruses

As anticipated, the fraction of HCCs attributable to HBV and HCV greatly varied by sex and age group, where 3.4% of HCCs diagnosed among females aged 35–39 years were attributable to HBV/HCV, compared to 63.7% among males aged 60–64 years. Since the same pooled ORs for each HBV and HCV were applied across sex and age groups, this difference is due to the prevalence of these viruses in different sex and age-groups.

Numerous studies use NHANES data to estimate the prevalence of HBV and HCV in the US.[238-241] Yet, some groups with the highest burden of these infections (in particular HCV), such as those incarcerated or experiencing homelessness, are excluded from the NHANES sampling frame.[242] Edlin and colleagues estimated, based on NHANES data collected from 2003–2010 and a systematic literature search for HCV RNA prevalence among NHANES excluded groups (e.g., incarcerated, homeless, hospitalized, nursing home residents, military personnel and those living on Indian reservations), that 0.8 million HCV RNA positive individuals were missing from the NHANES sampling frame, while 2.2 million individuals were HCV RNA positive and captured by the NHANES sampling frame.[242] This equates to 23% of the HCV RNA positive population being missed by NHANES. If we apply this 23% figure to our weighted versus weighted and imputed data, we see that the imputed data estimates are 23.4% higher than the non-imputed data for males (1.54% versus 1.18%) and 37.4% higher for females (0.70% versus 0.48%). Unlike prior analyses that excluded individuals who tested anti-HCV positive but were missing an HCV RNA result,[238,243] we utilized the anti-HCV result as a variable in the imputation model. In summary, while we did not directly account for those outside of the NHANES sampling frame, we produced HCV RNA estimates comparable to those reported by authors who used data on groups excluded from NHANES to readjust NHANES estimates. Finally, as we performed the imputations, we were able to retain the sex and age-group granularity desired (the Edlin et al. estimate is for the entire US population).

The introduction of highly effective direct-acting antiviral agents in late 2014,[10] and their subsequent uptake, could have weakened the relationship between HCV and HCC by 2017 and decreased the prevalence of chronic HCV infection. Yet our analysis, which used case-control studies published from 1991–2009 and utilized HCV prevalence data collected from 1999–2010, could not

account for the possible impact of curative HCV treatment. The extent to which HCV treatment was widely available and that treatment received from 2014–2016 would reduce HCC incidence in 2017 is unclear (though, we recognize that highly effective direct-acting antiviral agents are effective in patients with advanced disease and can therefore impact short-term HCC risk). However, given that there were ~2 years between their introduction and 2017, we believe the possible impact would be minor. Furthermore, therapy for the treatment of chronic HBV infection, though not curative, reduces the risk of HCC by decreasing liver inflammation and supressing viral replication.[244] The more effective treatments for chronic HBV infection were introduced in the mid-2000s and have the potential to weaken the relationship between HBV and HCC; however, our analyses could not directly account for the possible effect of chronic HBV treatment on HCC risk.[244]

4.2. Epstein-Barr virus

By including EBV's role in childhood cancers, DLBCL, and gastric carcinoma, associations traditionally not included in PAF analyses, an additional ~4250 cancers were attributable to EBV. The inclusion of DLBCL and gastric carcinoma altered the distribution of infection-attributable cancers; EBV comprised 11.1% of the infection-attributable cancer burden in adults but if DLBCL and gastric cancer were omitted, EBV would have instead been responsible for 5.9% of the infection-attributable cancer burden. Hence, there may be more EBV-attributable cases than generally recognized. EBV, often acquired early in life, establishes lifelong latency in more than 90% of adults worldwide.[34] Despite promising efforts to develop a vaccine, there is currently no way to prevent EBV infection.[245,246]

4.3. Human papillomavirus

For anal SCC, our finding that high-risk HPV DNA prevalence is higher in women than men is consistent with larger case series, such as the one performed by Frisch and colleagues with 386 anal cancer cases in Denmark and Sweden (95% in women versus 83% in men).[247] The result that more than 90% of women and 100% of PLWH with anal SCC tested positive for at least one HR-HPV is comparable to the proportion of cervical cancer testing HPV positive.[248] HPV's role in cervical cancer was responsible for 5.1% of all cancers among younger women (aged 20–49). While HPV vaccination efforts have been underway since the Food and Drug Agency approved the first HPV vaccine in 2006, females vaccinated at age 11-12 in 2006, would be 22 years old at the most in 2017. For this reason, the burden of cervical cancers remains high in 2017, but will presumably decrease in subsequent years.

While HPV-prevention efforts often focus on women, we found that of all cancers diagnosed in males in 2017, 1.8% (n = 14,754) were due to HPV. HPV prevalence in HNCs is higher in North America compared to other continents.[117] We included only those studies utilizing the gold standard detection method (E6/E7 mRNA detection) and considered HPV16 prevalence only; however, it is possible that other HPV types play a role in the carcinogenesis of HNCs. By restricting to HPV16 (which is by far the most prevalent HPV type found in HNCs),[117] a small proportion of cancers may have been missed.

4.4. Strengths and limitations

We sought to provide comprehensive estimates of the role of infections in cancer incidence in the US by (i) correcting for measurement error (*H. pylori* and NCGC association), (ii) imputing missing data (HBV, HCV and *H. pylori* population prevalence estimates), (iii) including additional infection-related cancers (intra and extrahepatic bile duct cancer [HBV and HCV], esophageal adenocarcinoma [*H. pylori*], DLBCL and gastric cancer [EBV]), (iv) utilizing different PAF estimates for PLWH where the data permitted (Hodgkin lymphoma and DLBCL [EBV]), and (v) defining the distribution of infection-attributable cancers for children, by sex, and for younger versus older adults. However, several limitations related to these estimates need to be mentioned. First, some PAF inputs were based on sparse data, in particular, for HBV and extrahepatic bile duct cancer (2 studies),[249,250] *H. pylori* and gastric mucosa-associated lymphoid tissue and DLBCL (one study),[122] EBV and Burkitt lymphoma in adults (one study),[140] and HPV in vaginal cancer (two studies).[171,172] We caution readers in interpreting the estimates for these three cancers. Second, in addition to having limited data for some cancers, several PAF calculations were based on data published more than 20 years ago; specifically, those used to estimate the measures of association for HBV, HCV, and *H. pylori*. Third, while most studies originated in the US, for certain infection-cancer pairs, we had to rely on studies conducted in Canada and Europe. Fourth, PAF equation 1 (used for HBV, HCV, and *H. pylori*) assumes no confounding between the exposure and the disease.[69] While we selected studies that matched and/or adjusted on strong confounders, residual confounding cannot be ruled out entirely. While residual confounding cannot explain the strong associations between infections and cancers included in this analysis, it could have a minor impact on the magnitude of those associations and the resulting PAFs. For multiple infections causing the same cancers (other than HBV/HCV and HCC), we made the simplifying assumption that the infections were independent causes of the given cancer; however, this

assumption may not account for potential interactions between infections. While not a limitation *per se*, we caution against extrapolating the estimates provided here to other regions. Finally, we remind readers that the burden of infection-attributable cancers is higher in certain groups, such as PLWH, organ transplant recipients, Indigenous peoples, immigrants, people who inject drugs, and the incarcerated. Future research focusing on populations expected to have a higher burden of infection-attributable cancers is useful for helping to implement prevention measures.

5. CONCLUSION

Infections were estimated to be responsible for 4.3% of cancers diagnosed in the US in 2017 and therefore represent an important target for the development of prevention efforts (for EBV) and continuation of current approaches (for HBV, HCV, *H. pylori* and HPV) to reduce their prevalence and associated disease burden.

Table 1. Overview of carcinogenic infections, included cancers, and criteria for selecting PAF inputs

Infection and cancers included (ICD-O-3 codes)	Inclusion criteria	
PAF calculated via formula 1 – prevalence of infection in the general population	**Prevalence estimates**	**Measure of association**
Hepatitis B virus Hepatocellular carcinoma (8170–8175) Intrahepatic bile duct (C22.1) Extrahepatic bile duct (C24.0)	Population-based serosurvey HBsAg detection Age and sex data	
Hepatitis C virus Hepatocellular carcinoma (8170–8175) Intrahepatic bile duct (C22.1) Extrahepatic bile duct (C24.0) Burkitt lymphoma (9687) Chronic lymphocytic leukemia/small lymphocytic lymphoma (9823) Diffuse large B-cell lymphoma, NOS (9680) Lymphoplasmacytic lymphoma (9671) Marginal zone B-cell lymphoma, NOS (9699)	Population-based serosurvey HCV RNA detection Age and sex data	Hepatitis infection confirmed by serology (HBsAg, anti-HCV [with or without Recombinant ImmunoBlot Assay confirmation], HCV RNA), 10 or more cases, US-based study population Hepatocellular carcinoma: controls without liver disease
Helicobacter pylori NCGC (C16.1-16.6 + proportion of C16.8, C16.9)[a] Gastric mucosal-associated lymphoid tissue lymphoma & diffuse large B-cell lymphoma (C16.1-16.9; 9699 & 9680)[b] Esophageal adenocarcinoma (C15; 8050-8083)	Population-based serosurvey ELISA detection Age and sex data	*H. pylori* infection confirmed by serology (ELISA, EIA or immunoblot), 10 or more cancer cases, North American, European or Australian and New Zealand study populations NCGC: prospective serology collection (~10 years before diagnosis), data required to correct sensitivity and specificity (if ELISA or EIA) Esophageal adenocarcinoma only: cohort, nested case-control or case-control studies, controls without gastrointestinal symptoms and not undergoing endoscopy for purposes other than screening, US based study population
PAF calculated via formula 3 – prevalence of infection in cancer tissue		
Epstein-Barr virus Burkitt lymphoma (9687) Hodgkin lymphoma (C81) Extranodal NK/T-cell lymphoma-nasal type (9719) Nasopharyngeal carcinoma (C11.0-9, 8020-21, 8070-73, 8082-83) Gastric carcinoma Diffuse large B-cell lymphoma, NOS (9680)	EBV detected via EBER ISH and for Hodgkin lymphoma via EBER ISH or LMP-1 in cancer tissues collected from North American based study populations or for Hodgkin lymphoma among PLWH in North American or European based study populations At least five cases if study population is HIV+, eight cases if pediatric population, and 10 cases if general adult population Hodgkin lymphoma: EBV prevalence reported by age group (aged 0–9, 10–19, 20–44, ≥45 years) Gastric carcinoma: EBV prevalence estimates by sex	
Human papillomavirus Cervix (C53) Anus, SCC (C21.0-C.21.2, C21.8, 8050-8052, 8070-8076, 8083-8084, 8123-8124) Penis (C60) Vagina (C52) Vulva (C51) Oropharynx (C01.9, C02.8, C02.4, C05.1, C05.2, C14.2, C09, C10) Oral cavity (C00.4–0.5, C00.9, C02.0–C02.9, C03, C04, C05.0, C05.8, C05.9, C06, C14.8) Larynx (C32)	Necessary cause[5,26] At least 10 invasive, non-recurrent cancer tissue specimens arising from at least 10 cancer cases from North American study populations Polymerase chain reaction detection for HPV Published in or after 1995 (anogenital cancers) or published in or after 2000 (head and neck cancers) Anogenital cancers: detection and results presented for high-risk HPV types Anal cancer: HPV results for SCC histology Vulvar cancer: data stratified by age or available upon request Head and neck cancers: infection with genotype 16 via the detection of E6 and/or E7 oncoproteins, site specific results (e.g., base of tongue versus oral tongue)	
Human herpesvirus, type 8 Kaposi sarcoma (9140) Primary effusion lymphoma (9678)	Necessary cause[1] Universally associated with HHV-8[251]	
Human T-cell lymphotropic virus, type 1 Adult T-cell leukemia/lymphoma (9827)	Necessary cause[228]	
Human immunodeficiency virus, type 1 [causal infectious agent] Kaposi sarcoma [HHV-8] (9140)[c] Cervix [HPV] (C53)[c] Burkitt lymphoma [EBV] (9687) Diffuse large B-cell lymphoma, NOS [EBV] (9680) Hodgkin lymphoma [EBV] (C81) Anus SCC [HPV] (see HPV section)	In addition to the inclusion criteria already listed under the respective infection (EBV or HPV) and cancer, here the sample size requirement was five or more cancer from North America or Europe	

EBER ISH = EBV-encoded RNA *in situ* hybridization, EBV = Epstein-Barr virus, EIA = enzyme immunoassay, ELISA = enzyme-linked immunosorbent assay, HBsAg = hepatitis B virus surface antigen, HCV = hepatitis C virus, *H. pylori* = *Helicobacter pylori*, LMP-1 = latent membrane protein 1, NCGC = non-cardia gastric cancer, NK = natural killer, NOS = not otherwise specified, PAF = population attributable fraction, SCC = squamous cell carcinoma, US = United States

a. NCGC incidence counts were adjusted by reassigning a proportion of 'overlapping lesion' and 'NOS' GC to NCGC – see methods.

b. In ICD-O-3, this cancer's morphology is referred to as marginal zone lymphoma.

c. While Kaposi sarcoma (HHV-8) and cervical cancer (HPV) are related to HIV, they were not considered separately in this analysis.

Table 2. Infections where PAFs were estimated using the prevalence of the infection in the population and measures of association

Infection Cancer(s)	Data used to estimate measure of association	Pooled OR (95% CI)	Source of prevalence data	Range of prevalence estimates by age group (years), by sex
Helicobacter pylori				
Gastric, non-cardia	Pooled ORs from nested case-control studies from the US, Europe and Australia with fixed effects: five studies[130-132,252,253] that used ELISA or EIA corrected for measurement error, and three studies[106,107,133] that used immunoblot	12.8 (8.5–19.3)	One cycle of NHANES data collected 1999–2000	Males 14.6% (aged 10–14) to 51.0% (aged 80–84) Females 9.3% (aged 10–14) to 49.6% (aged 70–74)
Gastric, MALT & DLBCL	One study of 20 cases matched to 80 controls from the US[122]	7.9 (1.6–38.1)		
Esophageal adenocarcinoma (protective effect)	Pooled ORs from case-control and nested-case control studies from the US with fixed effects: three studies[254-256] used ELISA and one study used immunoblot[257]	0.73 (0.55–0.95)		
Hepatitis B virus				
Hepatocellular carcinoma	Pooled ORs from four case-control studies from the US[258-261]	24.2 (14.5–40.3)	Six cycles of NHANES data collected 1999–2010	Males 0.2% (aged 6–29) to 1.0% (aged 50–59) Females 0.1% (aged 6–29) to 0.4% (aged 30–39)
Intrahepatic bile duct	Pooled ORs from four case-control studies from the US[249,250,262,263]	3.4 (1.2–9.4)		
Extrahepatic bile duct	Pooled ORs from two case-control studies from the US[249,250]	2.4 (1.6–3.4)		
Hepatitis C virus				
Hepatocellular carcinoma	Pooled ORs from five case-control studies from the US[258,259,261,264,265]	29.8 (11.9–74.6)	Six cycles of NHANES data collected 1999–2010	Males 0.2% (aged 6–29) to 4.5% (aged 45–49) Females 0.1% (aged 6–29) to 2.5% (aged 45–49)
Intrahepatic bile duct	Pooled ORs from four case-control studies from the US[249,250,262,263]	4.5 (3.5–5.7)		
Extrahepatic bile duct	Pooled ORs from three case-control studies from the US[249,250,266]	3.1 (2.4–4.1)		
BL/L (aged ≥50 years only)	ORs from five studies (from Australia, Canada, Europe and the US) that assessed HCV seropositivity in the InterLymph Non-Hodgkin Lymphoma Subtypes Project	4.1 (1.1–15.4)		
CLL/SLL		2.08 (1.23–3.49)		
DLBCL		Males: 2.17 (1.44–1.18) Females: 1.98 (1.18–3.34)		
LPL		2.51 (1.03–6.17)		
MZL		3.04 (1.65–5.6)		

BL/L = Burkitt lymphoma/leukemia, CI = confidence interval, CLL/SLL = chronic lymphocytic leukemia/small lymphocytic lymphoma, DLBCL = diffuse large B-cell lymphoma, EIA = enzyme immunoassay, ELISA = enzyme linked immunosorbent assay, HCV = hepatitis C virus, LPL = lymphoplasmacytic lymphoma, MALT = mucosa-associated lymphoid tissue, MZL = marginal zone lymphoma, NHANES = National Health and Nutrition Examination Survey, OR = odds ratio, PAF = population attributable fraction, US = United States

Table 3. Infections where PAFs were estimated using the prevalence of infection in cancer tissue

Infection Cancer	Source of prevalence estimates[a]	N cases used to estimate PAF	Sex/age/HIV group (age, in years)	PAF % (95% CI)
Epstein-Barr virus				
Burkitt lymphoma	7 studies	397	0–19	15.5 (8.1–23.0)
	1 study	51	≥20	Varied by age[b]
Hodgkin lymphoma	4 studies	148	0–9	62.2 (41.8–82.5)
	7 studies	443	10–19	22.3 (13.3–32.7)
	4 studies	983	15–44	20.5 (18.0–23.1)
	3 studies	369	≥45	42.5 (33.0–52.1)
	6 studies	282	HIV+ adults	92.9 (89.9–95.9)
Nasopharyngeal carcinoma	2 studies	16	0–19	100.0 (63.1–100.0)
	7 studies	629	≥20	61.2 (45.1–77.2)
ENKTL, nasal type	Part of diagnostic criteria	NA	≥20	100.0
DLBCL	14 studies[c]	5164	HIV- adults	4.9 (3.3–6.5)
	5 studies	264	HIV+ adults	45.7 (27.9–63.6)
Gastric cancer	7 studies	541	Male	13.6 (8.7–19.3)
		321	Female	1.9 (0.3–4.2)
Human papillomavirus, high-risk types,[d] anogenital tract cancers				
Anal SCC	5 studies	175	Male	90.2 (80.2–97.3)
		260	Female	96.3 (90.0–99.8)
Cervix	Necessary cause	NA	Female	100.0
Penis	5 studies	269	Male	38.6 (17.9–59.4)
Vagina	2 studies	85	Female	72.2 (62.8–81.7)
Vulva	6 studies	53	<50	74.4 (62.7–86.0)
		230	≥50	45.7 (21.9–69.4)
Human papillomavirus, type 16, head and neck cancers				
Oropharynx	17 studies	1905	≥20	60.3 (51.2–69.1)
Oral cavity	7 studies	683	≥20	7.9 (3.3–14.0)
Larynx	5 studies	194	≥20	12.7 (3.7–25.4)
Human herpesvirus, type 8				
Kaposi sarcoma	Necessary cause	NA	≥20	100.0
Primary effusion lymphoma	Universally associated with HHV-8	NA	≥20	100.0
Human T-cell lymphotropic virus, type 1				
Adult T-cell leukemia and lymphoma	Necessary cause	NA	≥20	100.0

CI = confidence interval, DLBCL = diffuse large B-cell lymphoma, ENKTL = extranodal natural killer T-cell lymphoma, HHV-8 = human herpesvirus type 8, HIV = human immunodeficiency virus, NA = not applicable, PAF = population attributable fraction, SCC = squamous cell carcinoma

a. The characteristics of included studies are reported in the supplement under their respective infection and cancers.

b. The prevalence estimates of 55% (aged 20–34), 33% (aged 35–59), 25% (aged ≥60) were used.

c. Included 13 studies where study authors reported that the study population was HIV negative and/or immunocompetent, and one study[267] with 567 DLBCL cases that did not report HIV status.

d. High-risk human papillomavirus (HPV) types include types classified by the International Agency for Research on Cancer as Group 1 (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 & 59), Group 2A (68) and Group 2B (34, 53, 66, 70 & 73) carcinogens. HPV97 was also considered a high-risk type.

Table 4. Estimates of the number and percent of cancers attributable to infections among individuals aged ≥20 years in the United States in 2017

Infection Cancer(s)	Aged ≥20 years			Aged 20–49 years				Aged ≥50 years			
	All			Males		Females		Males		Females	
	N	AC	%[a] Attributable (95% CI)	N	%[a] Attributable (95% CI)	N	%[a] Attributable (95% CI)	N	%[a] Attributable (95% CI)	N	%[a] Attributable (95% CI)
Hepatitis B virus											
Hepatocellular carcinoma[b]	24,190	2269	9.4 (2.9–14.6)	754	9.4 (2.3–15.1)	241	5.2 (1.0–8.9)	17,822	10.5 (3.6–16.0)	5,373	5.7 (0.8–10.0)
Intrahepatic bile duct[b]	5590	46	0.8 (0.1–1.5)	179	0.9 (0.2–1.7)	196	0.6 (0.1–1.1)	2726	1.1 (0.1–1.0)	2486	0.6 (0.1–1.1)
Extrahepatic bile duct[b]	3505	19	0.5 (0.1–1.0)	119	0.7 (0.1–1.3)	69	<0.1 (0.0–0.7)	1850	0.7 (0.1–1.2)	1467	0.4 (0.0–0.7)
Hepatitis C virus											
Hepatocellular carcinoma[b]	24,190	7764	32.1 (9.2–45.0)	754	18.8 (1.7–30.3)	241	11.3 (0.0–19.8)	17,822	36.8 (11.8–50.1)	5373	19.1 (1.8–31.2)
Intrahepatic bile duct[b]	5590	272	4.9 (0.5–8.9)	179	2.7 (0.1–5.2)	196	1.9 (0.0–3.8)	2726	7.0 (0.8–12.5)	2489	3.0 (0.1–5.7)
Extrahepatic bile duct[b]	3505	99	2.8 (1.2–4.4)	119	1.9 (0.7–3.2)	69	1.2 (0.2–2.2)	1850	3.9 (1.7–5.9)	1467	1.7 (0.5–2.8)
B-cell NHLs[c]	52,082	980	1.9 (0.4–3.4)	2,828	0.9 (0.2–1.6)	1876	0.6 (0.0–1.2)	29,960	2.5 (0.6–4.4)	18,013	1.0 (0.0–2.0)
Helicobacter pylori											
Gastric, non-cardia	14,539	11,766	80.6 (70.7–87.6)	788	75.8 (64.4–83.6)	932	73.8 (60.8–82.5)	6748	81.3 (70.9–88.0)	6072	82.2 (72.7–88.4)
Gastric, MALT & DLBCL	1950	1380	70.8 (0.8–90.5)	102	52.6 (0.0–70.4)	80	56.9 (0.0–77.7)	983	71.8 (0.1–92.0)	785	73.3 (1.9–92.6)
Esophageal ADC[d]	11,611	*-1266*	*10.6 (1.8–20.8)*	534	*8.5 (1.7–15.8)*	91	*7.6 (1.4–14.2)*	9423	*20.6 (1.8–20.6)*	1563	*12.6 (2.2–24.2)*
Epstein-Barr virus[c]											
Burkitt lymphoma	989	334	33.8 (19.9–58.1)	300	43.1 (22.6–64.9)	94	44.7 (23.3–66.8)	397	27.7 (18.0–53.3)	198	26.9 (18.0–53.3)
Hodgkin lymphoma[e]	7580	2510	33.1 (27.1–39.2)	2277	26.9 (23.4–30.3)	1967	22.8 (19.6–26.1)	1907	44.1 (34.7–53.5)	1429	42.5 (33.0–52.1)
NPC	1602	980	61.2 (45.1–77.2)	328	61.2 (45.1–77.2)	99	61.2 (45.1–77.2)	851	61.2 (45.1–77.2)	324	61.2 (45.1–77.2)
ENKTL – nasal type	186	186	100.0	43	100.0	17	100.0	82	100.0	44	100.0
DLBCL[e]	27,032	1571	5.8 (3.8–7.8)	1880	10.9 (6.9–14.9)	1263	4.9 (3.3–6.5)	13,271	5.9 (3.9–8.0)	10,618	4.9 (3.3–6.5)
Gastric	26,248	2,361	9.0 (5.4–13.4)	1445	13.6 (8.6–19.3)	1258	1.9 (0.3–4.2)	14,571	13.6 (8.6–19.3)	8974	1.9 (0.3–4.2)

AC = attributable cases, ADC = adenocarcinoma, ATLL = adult T-cell leukemia/lymphoma, CI = confidence interval, DLBCL = diffuse large B-cell lymphoma, EBV = Epstein Barr virus, ENKTL = extranodal natural killer T-cell lymphoma, HBV = hepatitis B virus, HCV = hepatitis C virus, HHV-8 = human herpesvirus type 8, HPV = human papillomavirus, HTLV-1 = human T-cell lymphotropic virus type 1, *H. pylori* = *Helicobacter pylori*, MALT = mucosa-associated lymphoid tissue, NA = not applicable, NHL = non-Hodgkin lymphoma, N = number of incident cancers diagnosed in 2017, NPC = Nasopharyngeal carcinoma, PAF = population attributable fraction, PEL = primary effusion lymphoma, SCC = squamous cell carcinoma

a. The % attributable was calculated by dividing the number of cases attributable to infection by the number of the associated cancer cases. It differs from PAF for which some cancers varied by sex and/or age.

b. The PAFs for each HBV and HCV for HCC, intrahepatic and extrahepatic bile duct were combined, but the individual % attributable are displayed to show the individual impact of the viruses.

c. PAFs were calculated for specific B-cell lymphomas (see Table 2 and Supplementary Table 6).

d. Due to *H. pylori's* protective effect in this cancer, cases attributable to esophageal adenocarcinoma were subtracted from the total cases attributable to *H. pylori.*

e. Accounts for differing EBV-attributable proportions among the general population and the population living with HIV for Hodgkin lymphoma and DLBCL (see supplement for details).

Table 4. Estimates of the number and percent of cancers attributable to infections among individuals aged ≥20 years in the United States in 2017 (continued)

Infection Cancer(s)	Aged ≥20 years			Aged 20–49 years				Aged ≥50 years			
	All			Males		Females		Males		Females	
	N	AC	%[a] Attributable (95% CI)	N	%[a] Attributable (95% CI)	N	%[a] Attributable (95% CI)	N	%[a] Attributable (95% CI)	N	%[a] Attributable (95% CI)
HPV, high-risk types											
Anus SCC	6451	6086	94.3 (86.8–99.0)	361	90.2 (80.2–97.3)	453	96.3 (90.0–99.8)	1746	90.2 (80.2–97.3)	3891	96.3 (90.0–99.8)
Cervix	12,829	12,829	100.0	NA		6199	100.0	NA		6630	100.0
Penis	1480	572	38.6 (17.9–59.4)	142	38.6 (17.9–59.4)	NA		1,338	38.6 (17.9–59.4)	NA	
Vagina	1335	964	72.2 (62.8–81.7)	NA		142	72.2 (62.8–81.7)	NA		1193	72.2 (62.8–81.7)
Vulva	5408	2659	49.2 (26.9–71.4)	NA		659	74.4 (62.7–86.0)	NA		4749	45.7 (21.9–69.4)
HPV, type 16											
Oropharynx	20,892	12,599	60.3 (51.2–69.1)	1575	60.3 (51.2–69.1)	363	60.3 (51.2–69.1)	15,554	60.3 (51.2–69.1)	3400	60.3 (51.2–69.1)
Oral cavity	15,269	1212	7.9 (3.3–14.0)	1046	7.9 (3.3–14.0)	696	7.9 (3.3–14.0)	8075	7.9 (3.3–14.0)	5452	7.9 (3.3–14.0)
Larynx	12,154	1547	12.7 (3.7–25.4)	516	12.7 (3.7–25.4)	210	12.7 (3.7–25.4)	9123	12.7 (3.7–25.4)	2305	12.7 (3.7–25.4)
HHV-8											
Kaposi sarcoma	1018	1018	100.0	503	100.0	28	100.0	403	100.0	84	100.0
PEL	51	51	100.0	20	100.0	<16	100.0	31	100.0	<16	100.0
HTLV-1											
ATLL	659	659	100.0	204	100.0	102	100.0	204	100.0	149	100.0
Overall											
All associated cancers[f]	213,079	71,469	33.5 (25.3–48.2)	13,275	33.5 (24.4–41.7)	14,841	61.3 (56.8–64.0)	116,648	28.2 (19.0–47.2)	68,314	36.6 (28.4–46.6)
All cancers	**1,666,102**	**71,469**	**4.3 (3.2–6.2)**	**68,632**	**6.5 (4.7–8.1)**	**122,827**	**7.4 (6.9–7.7)**	**773,672**	**4.3 (2.9–7.1)**	**700,971**	**3.6 (2.8–4.5)**

AC = attributable cases, ADC = adenocarcinoma, ATLL = adult T-cell leukemia/lymphoma, CI = confidence interval, DLBCL = diffuse large B-cell lymphoma, EBV = Epstein Barr virus, ENKTL = extranodal natural killer T-cell lymphoma, HBV = hepatitis B virus, HCV = hepatitis C virus, HHV-8 = human herpesvirus type 8, HPV = human papillomavirus, HTLV-1 = human T-cell lymphotropic virus type 1, *H. pylori* = *Helicobacter pylori*, MALT = mucosa-associated lymphoid tissue, NA = not applicable, NHL = non-Hodgkin lymphoma, N = number of incident cancers diagnosed in 2017, NPC = Nasopharyngeal carcinoma, PAF = population attributable fraction, PEL = primary effusion lymphoma, SCC = squamous cell cancer

a. The % attributable was calculated by dividing the number of cases attributable to infection by the number of the associated cancer cases. It differs from PAF for which some cancers varied by sex and/or age.

b. The PAFs for each HBV and HCV for HCC, intrahepatic and extrahepatic bile duct were combined, but the individual % attributable are displayed to show the individual impact of the viruses.

c. PAFs were calculated for specific B-cell lymphomas (see Table 2 and Supplementary Table 6).

d. Due to *H. pylori's* protective effect in this cancer, cases attributable to esophageal adenocarcinoma were subtracted from the total cases attributable to *H. pylori.*

e. Accounts for differing EBV-attributable proportions among the general population and the population living with HIV for Hodgkin lymphoma and DLBCL (see supplement for details).

f. Among cancers related to infections that appear in this table.

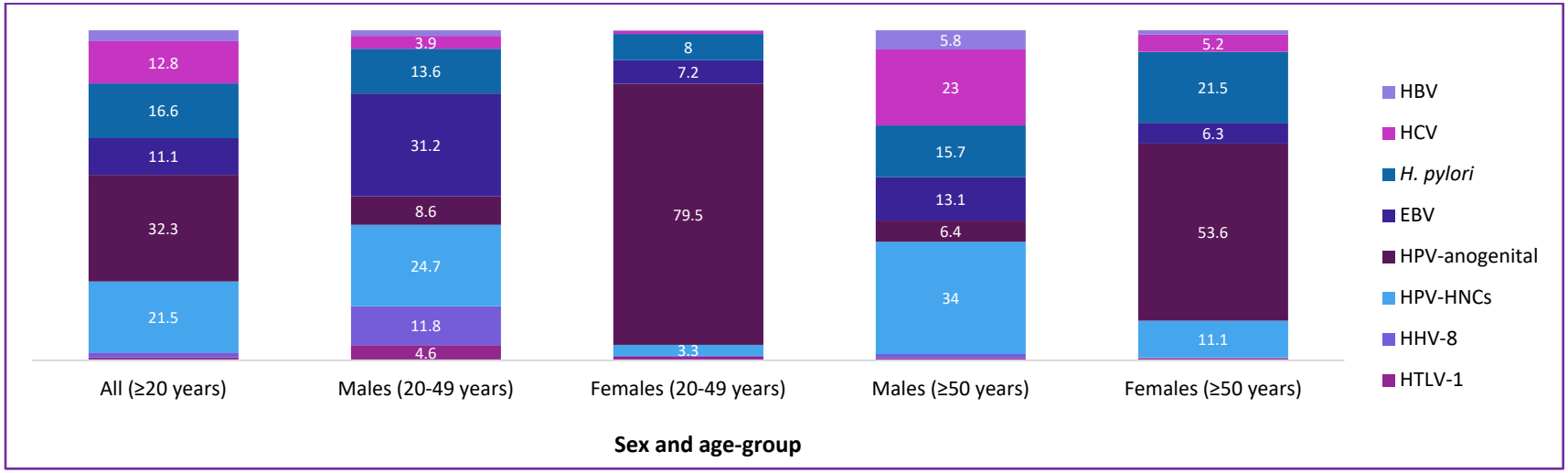
Table 5. Estimates of the percent and number of cancers attributable to EBV among individuals aged 0 to 19 years in the United States in 2017

Sex and age-group	All cancers			Burkitt lymphoma		Hodgkin lymphoma		Nasopharyngeal carcinoma[a]	
	N	AC	PAF % (95% CI)	N	AC	N	AC	N	AC
Males									
0–9 years	3836	46	1.2 (0.8–1.6)	68	11	57	35	…	…
10–19 years	4172	130	3.1 (1.5–3.8)	76	12	427	95	…	…
Total	8008	176	2.2 (1.3–3.1)	144	22	484	131	25	25
Females									
0–9 years	3236	17	0.5 (0.3–0.7)	25	4	21	13	…	…
10–19 years	3808	131	3.4 (1.7–4.3)	28	4	477	107	…	…
Total	7044	148	2.1 (1.3–3.0)	53	8	498	120	30	30
Overall	**15,052**	**324**	**2.2 (1.3–3.0)**	**197**	**31**	**982**	**250**	**43**	**43**

AC = attributable cases, CI = confidence interval, EBV = Epstein Barr virus, N = number of incident cancers diagnosed in 2017, PAF = population attributable fraction

… Aggregated due to small/suppressed cell counts.

Fig. 1. Distribution (%) of infection-attributable cancers in the US in 2017, overall and by sex and age group



EBV = Epstein-Barr virus, HBV = Hepatitis B virus, HCV = Hepatitis C virus, HHV-8 = human herpesvirus type 8, HNCs = head and neck cancers, HPV = human papillomavirus, HTLV-1 = human T-cell lymphotropic virus type 1, *H. pylori* = *Helicobacter pylori*

Supplementary material to manuscript #2

Supplementary material notes

- All confidence intervals (CIs) reported in this supplement are **95% CIs**.
- The meta-analyses are from a random effects model, unless otherwise specified in the title of the forest plot.

Acknowledgements

We are grateful to the following individuals for sharing unpublished data with us:

Dr. Jon H. Chung (Clinical development, Foundation Medicine, Cambridge, Massachusetts); **Dr. Janet Daling** (Program in Epidemiology, Fred Hutchinson Cancer Research Center Seattle, Washington); **Dr. Cecilie Dupont Harwood** (Department of Otorhinolaryngology, Head and Neck Surgery, and Audiology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark); **Dr. Terence Dunn** (College of Medicine, The University of Oklahoma, Norman, Oklahoma); **Dr. Julia Gargano** (Division of Viral Diseases, HPV Team, Centers for Disease Control and Prevention, Atlanta, Georgia); **Dr. Michael Herfs** (Laboratory of Experimental Pathology, GIGA-Cancer, University of Liège, Liège, Belgium); **Ms. Elysha Kolitz** (The University of Texas Southwestern Medical Center, Dallas, Texas); **Dr. Christina S. Kong** (Stanford Cancer Center & Lucile Packard Children's Hospital, Palo Alta, California); **Dr. Sam Mbulaiteye** (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland); **Dr. Margaret Madeleine** (Program in Epidemiology, Fred Hutchinson Cancer Research Center, Seattle, Washington); **Dr. Edyta C. Pirog** (Weill Cornell Medicine, Cornell University, Ithaca, New York).

Literature search

The search shown in **Table S1** was designed to capture knowledge syntheses (i.e., systematic reviews with or without meta-analyses, scoping reviews, etc.). This search captured 3,230 records of which 348 underwent full-text review.

Table S1. Search performed in MEDLINE(R) 1946–September 15, 2021

Individual infections	1.	exp Hepatitis B virus/ or exp Hepatitis B/ or exp Hepatitis C/ or exp Hepacivirus/ or (hepatitis virus* or hepatitis B or hepatitis C or HBV or HCV or hep B or hep C).tw,kf.
	2.	exp Herpesvirus 4, Human/ or exp Epstein-Barr Virus Infections/ or (herpesvirus type 4 or herpesvirus 4 or ebv or hhv4 or hhv-4).tw,kf. or ((epstein-Barr or epstein Barr) adj2 (virus* or viral*)).tw,kf.
	3.	exp HTLV-I Infections/ or exp Human T-lymphotropic virus 1/ or (human T-cell lymphotropic virus or Human T-lymphotropic virus or HTLV-1 or HTLV1).tw,kf.
	4.	exp Herpesvirus 8, Human/ or (human herpesvirus 8 or human herpesvirus type 8 or sarcoma-associated herpesvirus or Kaposi sarcoma-associated herpesvirus or HHV-8 or HHV8 or KSHV).tw,kf. or (Kaposi* adj3 (virus* or viral*)).tw,kf.
	5.	exp Helicobacter/ or exp Helicobacter infection/ or (helicobacter or pylori or pyloridis or HP or campylobacter, H* pylori).tw,kf.
	6.	exp Papillomavirus Infections/ or exp Papillomaviridae/ or (human papillomavirus* or human papilloma virus* or hpv).tw,kf.
	7.	exp HIV Infections/ or exp HIV/ or (hiv or hiv-1 or hiv-2 or hiv1 or hiv2 or hiv infect* or deficiency virus).tw,kf. or (human immun* adj2 (virus* or viral*)).tw,kf.
	8.	exp Merkel cell polyomavirus/ or (merkel cell polyomavirus or MCV or MCPyV).tw,kf. or (merkel adj3 polyomavirus).tw,kf.
	9.	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
Cancer	10.	exp Neoplasms/ or (cancer* or neoplas* or tumor* or tumour* or malignan* or carcinoma* or metasta* or oncolog* or leukemi* or leukaemi* or lymphoma* or myeloma* or sarcoma* or squamous cell* or adenocarcinoma*).tw,kf.
Knowledge syntheses[a]	11.	(meta-analysis or systematic review).pt. or meta-analysis/ or systematic review/ or exp meta-analysis as topic/ or ((systematic* adj3 (review* or overview*)) or (methodologic* adj3 (review* or overview*))).ti,ab,kf. or ((quantitative adj3 (review* or overview* or synthes*)) or (research adj3 (integrati* or overview*))).ti,ab,kf. or ((integrative adj3 (review* or overview*)) or (collaborative adj3 (review* or overview*)) or (pool* adj3 analy*)).ti,ab,kf. or (data synthes* or data extraction* or data abstraction*).ti,ab,kf. or (handsearch* or hand search*).ti,ab,kf. or (met analy* or metanaly*).ti,ab,kf. or (meta regression* or metaregression*).ti,ab,kf. or (meta-analy* or metaanaly* or systematic review*).mp,hw. or (medline or cochrane or pubmed or medlars or embase or cinahl).ti,ab,hw. or (mantel haenszel or peto or der simonian or dersimonian or fixed effect* or latin square*).ti,ab,kf.
Limits	12.	9 and 10 and 11
	13.	limit 12 to English
	14.	limit 13 to humans

[a] The knowledge syntheses search terms were adapted from the Canadian Agency for Drugs and Technologies in Health (CADTH) database search filters. Ottawa: CADTH; 2016. [Available from: /resources/finding-evidence]

HEPATITIS B AND C VIRUSES (HBV, HCV)

Tables S2 and S3 display the general population prevalence estimates for the hepatitis viruses.

Table S2. Estimated prevalence of chronic HBV in the US, NHANES data collected 1999–2010

Sex-age group (years)	Sample				Weighted			Imputed + Weighted	
	Pos no.	Pos %	RSE[a] %	Missing[b] %	Pos %	RSE[a] %	Missing[b] %	Pos %	RSE[a] %
Males									
6–29	14	0.13	26.7	11.1	0.22	31.5	11.1	0.24	29.3
30–39	13	0.58	27.7	6.6	0.51	32.3	5.4	0.56	31.8
40–49	12	0.50	28.8	4.7	0.26	30.3	4.0	0.29	34.2
50–59	22	1.08	21.2	5.1	0.93	22.4	3.6	0.97	22.8
≥60	14	0.32	26.7	5.3	0.28	40.1	4.2	0.28	39.2
Overall	**75**	**0.35**	**11.5**	**8.2**	**0.38**	**14.7**	**7.0**	**0.41**	**7.2**
Females									
6–29	6	0.06	40.8	11.5	0.05	65.2	12.0	0.07	58.5
30–39	10	0.38	31.6	6.0	0.34	30.5	5.2	0.37	32.7
40–49	11	0.43	30.1	5.0	0.28	36.4	4.4	0.30	36.1
50–59	6	0.30	40.8	5.6	0.20	42.4	4.9	0.27	44.5
≥60	13	0.30	27.7	6.3	0.19	33.5	5.3	0.27	29.3
Overall	**46**	**0.21**	**14.7**	**8.7**	**0.18**	**18.7**	**7.5**	**0.22**	**5.7**

HBsAg = hepatitis B surface antigen, HBV = hepatitis B virus, NHANES = National Health and Nutrition Examination Survey, Pos = positive, RSE = relative standard error, US = United States

a. The RSE, which is calculated by dividing the estimate's standard error by the estimate itself, RSE's of <30% should be indicated in reporting.[230]

b. Missing refers to individuals who attended the interview and medical examination but do not have a test result for HBsAg infection.

Table S3. Estimated prevalence of chronic HCV in the US, NHANES data collected 1999–2010

Sex-age group (years)	Sample				Weighted			Imputed + Weighted	
	Pos no.	Pos %	RSE[a] %	Missing[b] %	Pos %	RSE[a] %	Missing[b] %	Pos %	RSE[a] %
Males									
6–29	7	0.07	37.8	11.1	0.16	44.9	11.2	0.19	42.3
30–34	7	0.65	37.7	7.5	0.43	33.5	6.6	0.78	36.4
35–39	18	1.54	23.4	6.4	1.19	23.6	5.2	1.28	22.9
40–44	34	2.78	16.9	6.1	2.67	18.8	5.8	3.42	18.1
45–49	48	4.21	14.1	5.6	3.83	17.3	4.3	4.45	16.6
50–54	45	3.86	14.6	6.6	2.87	17.7	5.5	4.33	17.6
55–59	20	2.38	22.1	6.5	1.62	31.6	4.4	2.14	26.7
60–64	26	2.16	19.4	5.7	1.40	26.5	3.7	1.53	24.5
≥65	15	0.48	25.8	5.8	0.27	29.8	4.8	0.51	24.2
Overall	**220**	**1.03**	**6.7**	**8.7**	**1.18**	**9.2**	**7.5**	**1.54**	**8.6**
Females									
6–29	6	0.06	40.8	11.6	0.06	59.9	12.0	0.07	54.9
30–34	5	0.37	44.6	7.0	0.43	53.9	6.3	0.51	47.3
35–39	12	0.95	28.7	5.5	0.75	34.3	4.8	0.93	34.0
40–44	14	1.08	26.6	5.8	0.83	30.2	5.6	1.22	29.1
45–49	25	2.06	19.8	6.2	1.58	22.2	5.1	2.47	21.3
50–54	19	1.71	22.7	7.2	1.05	30.9	6.5	1.60	28.6
55–59	7	0.81	37.6	5.4	0.35	38.7	4.5	0.67	43.1
60–64	8	0.62	35.2	5.9	0.46	45.3	4.5	0.54	38.1
≥65	11	0.41	27.3	7.1	0.22	48.4	5.9	0.34	31.2
Overall	**107**	**0.48**	**9.6**	**9.0**	**0.48**	**12.3**	**7.9**	**0.70**	**11.7**

HCV = hepatitis C virus, NHANES = National Health and Nutrition Examination Survey, Pos = positive, RSE = relative standard error, US = United States

a. The RSE, which is calculated by dividing the estimate's standard error by the estimate itself, RSE's of <30% should be indicated in reporting.[230]

b. Missing refers to individuals who attended the interview and medical examination but do not have a test result for HCV RNA infection (note, this can include those who tested anti-HCV positive but did not have sufficient volume of sera to be tested for HCV RNA).

Hepatocellular carcinoma (HCC)

Through inflammation of the liver (cirrhosis), HBV and HCV can cause the major liver cancer histological type – HCC.[1] Additionally, HBV can cause HCC directly through chromosomal integration.[268] Studies reporting a measure of association are detailed in **Table S4**.

Table S4. Characteristics of case-control studies conducted in the US on the association between HBV or HCV infection and hepatocellular carcinoma

Study[a]	Study population	Matching variables	Characteristics of participants	Detection method	Cases		Controls		OR (95% CI)[b]	Adjustment variables & remarks
					n/N	Pos %	n/N	Pos %		
Hassan 2009[265]	Cases: diagnosed HCC GI outpatient clinics at M.D. Anderson Cancer Center Controls: three healthy controls/case non-blood family members of patients recruited from radiology clinic; similar in age, sex, race/ethnicity, education level Recruited: 2000–2006 Diagnosed: 2000–2008	Sex, age group, race	Males: 245 cases; 615 controls Mean age (SE): 62 (0.7) for cases; 60 (NS) for controls	**Anti-HCV** (3rd gen. ELISA)	79/347	22.8	6/1075	0.6	79.2 (30.6–204.8)	Age, sex, race, educational level, cigarette smoking, alcohol consumption, diabetes mellitus, family history of cancer, HBsAg, anti-HBc
Ognjanovic 2009[264]	Cases: Los Angeles HCC Study (HCC cases were identified through the Los Angeles County Cancer Surveillance Program, a population-based cancer registry) Controls: two neighbourhood controls/case and from Health Care Financing Administration files Diagnosed: 1984–2001 Sera collection: 1992–NS	Sex, age (±5), race	Males: 82 cases; 139 controls Mean age (SD): 60.5 (10.3) for cases; 59.5 (10.7) for controls, range: 18–74 yrs	**Anti-HCV** via ELISA v2 kit, confirmed with RIBA	58/120	48.3	1/230	0.4	211.0 (40.01–4368)	None OR calculated in OpenEpi
Davila 2005[258]	Cases: HCC in SEER registries also enrolled in Medicare (aged ≥65 yrs) Controls: population-based non-cancer controls aged ≥65 yrs, matched 3:1 to cases on time of diagnosis Recruited: 1994-1999	Frequency matched	Males: 1352 cases; 2248 controls Minimum age for study: 65 Age ≥75: 1139 cases; 3260 controls	ICD-9 codes for **HBV**	182/ 2061	8.8	14/ 6183	0.2	23.94 (13.65-41.99)	Age, sex, race, SEER registry, Medicare/ Medicaid dual enrolment; HCV OR without diabetes
				ICD-09 codes for **HCV** or unspecified hepatitis diagnosed before 1992	406/ 2061	19.7	80/ 6183	1.3	24.42 (17.49-34.11)	

anti-HBc = total hepatitis B core antibody, CI = confidence interval, gen. = generation, GI = gastrointestinal, HBV = hepatitis B virus, HBsAg = hepatitis B surface antigen, HCC = hepatocellular carcinoma, HCV = hepatitis C virus, EIA = enzyme immunoassay, ELISA = enzyme-linked immunoassay, NS = not specified, OR = odds ratio, Pos = positive, RIBA = Recombinant ImmunoBlot Assay, SD = standard deviation, SE = standard error, SEER = Surveillance, Epidemiology, and End Results Program, US = United States, yrs = years

a. Inclusion criteria: hepatitis infection confirmed by serology (HBsAg, anti-HCV [with or without RIBA confirmation], HCV RNA), 10 or more HCC cases, controls without liver disease, US study population.

b. After the normalizing transformation is performed, the CIs listed in the table may not match those in the forest plot.

Table S4. Characteristics of case-control studies conducted in the US on the association between HBV or HCV infection and hepatocellular carcinoma (continued)

Study[a]	Study population	Matching variables	Characteristics of participants	Detection method	Cases		Controls		OR (95% CI)[b]	Adjustment variables & remarks
					n/N	Pos %	n/N	Pos %		
Hassan 2002[259]	Cases: HCC patients diagnosed at The University of Texas M. D. Anderson Cancer Center hospital Controls: histologically confirmed malignant neoplasms other than HCC, which included primary tumors of the GI tract (44.3%), urogenital tract (18.7%), respiratory tract (17.8%), and skin (19.1%) Diagnosed: 1994–1995	2 controls matched for sex, age (5 yrs), year of diagnosis to 1 case	Males: 87 cases; 174 controls Mean age (SD): 59.5 (10.7) for cases; 59.1 (10.9) for controls	**HBsAg** via ELISA	17/115	14.8	2/230	0.9	23.8 (3.9–141.6)	Alcohol consumption, cigarette smoking, diabetes mellitus, anti-HCV (for HBV only), HBsAg (for HCV only)
				Anti-HCV (2nd gen. ELISA) confirmed with RIBA	26/115	22.6	5/230	2.2	14.1 (4.0–49.7)	
Nomura 1996[260]	Nested case-control Cases: American men of Japanese ancestry with HCC, born between 1900-1919 living in Hawaii Controls: males without cancer selected from the cohort Recruited/diagnosed: NS	Age at examination, date of serum collection	All male	**HBsAg**	15/24	62.5	2/72	2.8	43.0 (5.7–325.5)	Not adjusted
				Anti-HCV via EIA and confirmed with RIBA (1st gen.)	0/23	0.0	0/67	0.0	Not computed due to a lack of exposure in cases and controls	
Di Bisceglie 1991[261]	Cases: consecutive HCC patients at Johns Hopkins Oncology Center Controls: patients with other malignant tumors (20% GI tract, 34% respiratory tract, 20% urogenital tract, and 16% breast, 10% neurological or hematological) at same institution Diagnosed: 1987–1988	Not matched	Males: 67 cases; 53 controls Mean (range) age at diagnosis: 52 (10–86) for cases; 55 (18–70) for controls	**HBsAg**	7/99	7.1	0/98 (0.5 added to empty cell)	0.0	11.31 (1.39–335.3)	Not adjusted; 0.5 added to empty cell (HBsAg controls)
				Anti-HCV	13/99	13.1	2/98	2.0	7.20 (1.78–48.22)	

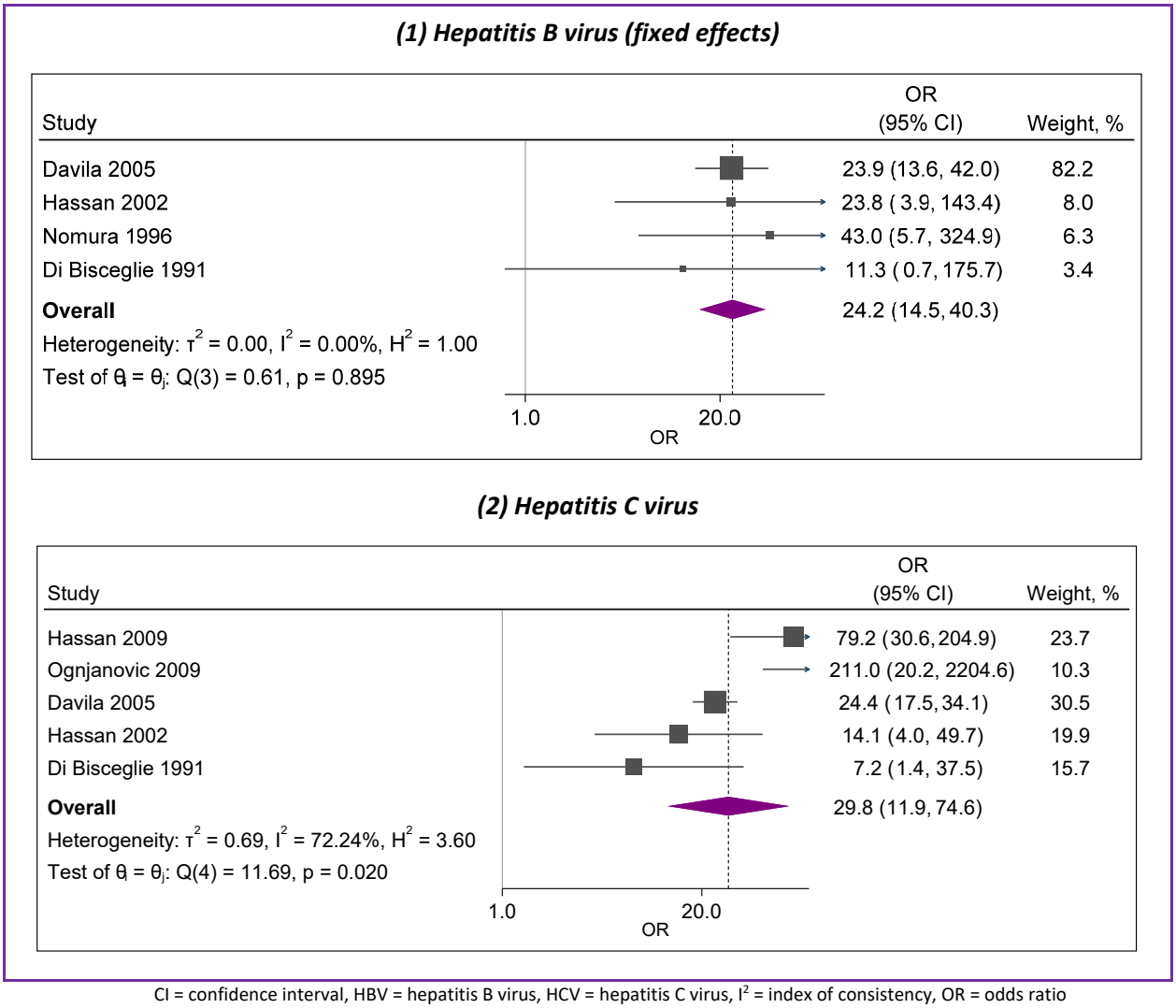
anti-HBc = total hepatitis B core antibody, CI = confidence interval, gen. = generation, GI = gastrointestinal, HBV = Hepatitis B virus, HBsAg = hepatitis B surface antigen, HCC = hepatocellular carcinoma, HCV = hepatitis C virus, EIA = enzyme immunoassay, ELISA = enzyme-linked immunoassay, NS = not specified, OR = odds ratio, Pos = positive, RIBA = Recombinant ImmunoBlot Assay, SD = standard deviation, SE = standard error, SEER = Surveillance, Epidemiology, and End Results Program, US = United States of America, yrs = years

a. Inclusion criteria: hepatitis infection confirmed by serology (HBsAg, anti-HCV [with or without RIBA confirmation], HCV RNA), 10 or more HCC cases, controls without liver disease, study population from the US.

b. After the normalizing transformation is performed, the CIs listed in the table may not match those in the forest plot.

Pooling four studies reporting on HBV and five studies reporting on HCV gave a pooled odds ratio (OR) of 24.2 (confidence interval [CI]: 14.5–40.3) for HBV and 29.8 (CI: 11.9–74.6) for HCV (**Fig. S1**).

Fig. S1. Pooled ORs for the association between each (1) HBV and (2) HCV and hepatocellular carcinoma



CI = confidence interval, HBV = hepatitis B virus, HCV = hepatitis C virus, I^2 = index of consistency, OR = odds ratio

The individual population attributable fractions (PAFs) for HBV ranged from 1.5–18.4% and for HCV from 1.9–56.2% (**Table S5**). After combining individual HBV and HCV PAFs estimates, 3.4–63.7% of HCCs were attributable to HBV and HCV.

Table S5. HBV and HCV associated PAFs (%) for hepatocellular carcinoma, by age group and sex

HCC sex-age group incidence (years)	HBV		HCV		Combined HBV-HCV PAF for 2017 %[a]	Partitioned PAFs [a]	
	Prevalence from age group (years)	Individual PAF %	Prevalence from age group (years)	Individual PAF %		HBV %	HCV %
Males							
20–24	6–29	5.3	6-29	5.1	10.1	5.2	4.9
25–29	6–29	5.3	6-29	5.1	10.1	5.2	4.9
30–34	6–29	5.3	6-29	5.1	10.1	5.2	4.9
35–39	6–29	5.3	6-29	5.1	10.1	5.2	4.9
40–44	30–39	11.6	30-34	18.4	27.9	10.7	17.1
45–49	30–39	11.6	35-39	26.9	35.3	10.6	24.7
50–54	40–49	6.2	40-44	49.7	52.8	5.9	46.9
55–59	40–49	6.2	45-49	56.2	59.0	5.9	53.1
60–64	50–59	18.4	50-54	55.6	63.7	15.9	47.9
65–69	50–59	18.4	55-59	38.2	49.5	16.1	33.4
70–74	≥60	6.2	60-64	30.6	34.9	5.9	29.0
≥75	≥60	6.2	≥60	12.7	18.1	5.9	12.2
Females							
20–24	6–29	1.5	6-29	1.9	3.4	1.5	1.9
25–29	6–29	1.5	6-29	1.9	3.4	1.5	1.9
30–34	6–29	1.5	6-29	1.9	3.4	1.5	1.9
35–39	6–29	1.5	6-29	1.9	3.4	1.5	1.9
40–44	30–39	7.9	30-34	12.9	19.8	7.5	12.3
45–49	30–39	7.9	35-39	21.2	27.4	7.4	20.0
50–54	40–49	6.6	40-44	26.0	30.9	6.2	24.7
55–59	40–49	6.6	45-49	41.6	45.5	6.2	39.2
60–64	50–59	5.8	50-54	31.6	35.6	5.5	30.1
65–69	50–59	5.8	55-59	16.2	21.0	5.6	15.5
70–74	≥60	5.9	60-64	13.4	18.5	5.7	12.9
≥75	≥60	5.9	≥60	9.0	14.4	5.7	8.7

HBV = hepatitis B virus, HCV = hepatitis C virus, HCC = hepatocellular carcinoma, PAF = population attributable fraction

[a.] The PAFs for HBV and HCV in HCC were combined with this equation: 1 – (1 – HBV PAF) * (1 – HCV PAF) then partitioned by determining the proportion of the summed number of attributable cases.[72]

Non-Hodgkin lymphoma (NHL)

Since NHLs are a heterogenous group of cancers and studies show that the magnitude of the association between HCV and NHL varies by subtype,[123,211,212] we utilized subtype specific measures of association. Data arising from the InterLymph Non-Hodgkin Lymphoma Subtypes Project, which pooled data from 11 mostly population-based case-control studies conducted in Australia, Europe and North America were used in the PAF calculations.[212,269] Of the 11 InterLymph studies, six[2] assessed HCV seropositivity via third-generation enzyme-linked

[2] The six studies were four population-based (region of recruitment and years cases diagnosed): British Columbia (Vancouver & Victoria, Canada; 2000–2004), UCSF1 (San Francisco, US; 1988–1995), SCALE (Denmark & Sweden, 1999–2002), New South Wales (Australian Capital Territory, 2000–2001); one mixed population-based and/or hospital-based: EpiLymph (Spain, France, Germany, Italy, Ireland, Czech Republic; 1998–2004; Italy and Germany were-population-based – the remainder were hospital-based), and one hospital-based: Italy – Aviano-Milan (1983–1992).

immunosorbent assay (ELISA).[269] The overall OR for the association for HCV and NHL was 1.81 (CI: 1.39–2.37). Subtypes that HCV demonstrated a statistically significant association with were included (**Table S6**). Notably, there were few cases of Burkitt lymphoma/leukemia (BL/L); however, since a similar magnitude of association (OR = 5.2, CI: 1.6–16.8) was also found in another large (33,940 NHL cases overall, 197 BL cases) study conducted in the US, we retained BL/L in the analysis.[123] An OR for HCV and BL/L in those aged <50 years was not calculated by the original study authors; we imputed 0.5 persons to the empty cell and calculated an OR of 1.47 (CI: 0.07–8.03). Since it was not statistically significant, we did not include BL/L among those <50 years in the PAF calculations. With only three cases of Waldenström's macroglobulinemia (the remaining 204 cases were lymphoplasmacytic lymphoma [LPL]), we applied the resulting PAF from LPL/WM to LPL incidence only.

Table S6. The association between HCV infection NHL subtypes as reported in the InterLymph study

NHL subtype	Cases		Controls		Adjusted OR (95% CI)	Adjustment variables
	n/N	Pos %	n/N	Pos %		
BL/L: age <50 years[270]	0/31	0.0	42/1933	2.2	--	Age, sex, race/ethnicity, study
BL/L: age ≥50 years[270]	3/33	9.1	109/4562	2.4	4.1 (1.1–15.4)	
CLL/SLL[271]	21/994	2.1	95/5354	1.8	2.08 (1.23–3.49)	
DLBCL[272]	63/1654	3.8	152/6898	2.2	Males: 2.17 (1.44–1.18) Females: 1.98 (1.18–3.34)	Age, sex, race/ethnicity, study, SES, history of autoimmune disease, any atopic disorder, blood transfusion, year of first OC use, age at first HT use, 1st degree family history – NHL, BMI as young adult, usual adult BMI, lifetime alcohol consumption, recreational sun exposure, field crop vegetable farmer, sewer & embroiderer, women's hairdresser, driver/material handling equipment operator
LPL/WM[273]	6/207[a]	2.9	95/5354	1.8	2.51 (1.03–6.17)	Age, sex, race/ethnicity, study, Sjögren syndrome, systemic lupus erythematosus, hay fever, usual adult weight, smoking duration, family history of hematological malignancy, medical occupation
MZL[274]	14/368	3.8	95/5354	1.8	3.04 (1.65–5.6)	Age, sex, race/ethnicity, study

BL/L = Burkitt lymphoma/leukemia, BMI = body mass index, CLL/SLL = chronic lymphocytic leukemia/small lymphocytic lymphoma, DLBCL = diffuse large B-cell lymphoma, HCV = hepatitis C virus, HT = hormone therapy, LPL/WM = lymphoplasmacytic lymphoma/Waldenström's macroglobulinemia, MZL = marginal zone lymphoma, NHL = non-Hodgkin lymphoma, Pos = positive, OC = oral contraceptive, SES = socioeconomic status

a. Among the 374 cases enrolled, only three were diagnosed with WM and the remainder LPL.

Intrahepatic and extrahepatic bile duct cancer

Pooling four studies reporting on intrahepatic bile duct cancer for each HBV and HCV (**Table S7**) gave a pooled OR of 3.4 for HBV and 4.5 for HCV (**Fig. S2**). Pooling two and three studies reporting on extrahepatic bile duct cancer for each HBV and HCV, respectively (Table S7) gave a pooled OR of 2.4 for HBV and 3.1 for HCV (**Fig. S3**).

Table S7. Characteristics of case-control studies conducted in the US on the association between HBV or HCV infection and intrahepatic and/or extrahepatic bile duct cancer

Study[a]	Study population	Matching variables	Characteristics of participants, age in years (SD)	Detection method	Cancer	Cases		Controls		OR[b] (95% CI)	Adjustment variables & remarks
						n/N	Pos %	n/N	Pos %		
Petrick 2017[249]	SEER-Medicare database Cases: Medicare beneficiaries enrolled continuously in Medicare Parts A and B for a minimum of three years prior to cancer diagnosis with ICC or ECC Controls: 5% random sample of Medicare-enrolled beneficiaries residing in the SEER 18 geographic regions and without prior cancer diagnoses Period: 2000–2011	None	>68 ICC cases: 78.0 (6.5) ECC cases: 79.2 (6.8) Controls: 76.6 (7.7)	ICD-9 codes for **HBV**	ICC	25/2092	1.2	1200/323,615	0.4	2.97 (1.97–4.46)	Age, race/ethnicity, geographic region, state buy-in status
					ECC	31/2981	1.0	1200/323,615	0.4	2.38 (1.65–3.44)	
				ICD-9 codes for **HCV**	ICC	58/2092	2.8	2161/323,615	0.7	4.67 (3.57–6.11)	
					ECC	57/2981	1.9	2161/323,615	0.7	3.18 (2.43–4.16)	
Choi 2016[262]	Cases: CCA patients seen at the Mayo Clinic from 2000–2014 Controls: recruited from the Mayo Clinic Biobank from 2009–2015, which comprises a collection of blood samples & health information from Mayo Clinic patients and other community volunteers (without a history of cancer other than nonmelanoma skin cancer)	Frequency-matched 1:2 for age (±5 yrs), sex, race, residence	Cases: 60.6 (13.1) Controls: 61.6 (13.5)	**HBsAg**	ICC	10/1169	0.9	8/4769	0.2	12.9 (2.69–61.61)	Propensity score adjustment: age, sex, race, obesity, etc.[c]
				Anti-HCV	ICC	23/1169	2.0	17/4769	0.4	1.95 (0.75–5.11)	
Welzel 2007[266]	SEER-Medicare database Cases: patients diagnosed with ECC or ICC enrolled in Medicare Parts A and B for at least three years before diagnosis Controls: individuals with no prior cancer diagnoses were selected from a 5% random sample of Medicare-enrolled beneficiaries who resided in the geographic regions of the SEER 11 registries Period: 1993–1999	Year of search for risk factors	>68 Cases: 78.7 (6.9)	ICD-9 codes for **HBV**	ECC	<5/549	<1.0	142/102,782	0.4	1.5 (0.2–11.0)	Age, gender, race, geographic location, state buy-in status

CI = confidence interval, ECC = extrahepatic cholangiocarcinoma, HBV = Hepatitis B virus, HBsAg = hepatitis B surface antigen, HCV = hepatitis C virus, ICC = intrahepatic cholangiocarcinoma, Pos = positive, SD = standard deviation, SEER = Surveillance, Epidemiology, and End Results Program

a. Inclusion criteria: hepatitis infection confirmed by serology (HBsAg, anti-HCV [with or without RIBA confirmation], HCV RNA), 10 or more ICC or ECC cases, study population from the US

b. After the normalizing transformation is performed, the CIs listed in the table may not match those in the forest plot.

c. As well as hypertension, diabetes, cerebrovascular accident, coronary artery disease, peripheral vascular disease, atrial fibrillation, non-alcoholic fatty liver disease, non-alcoholic steatohepatitis, primary sclerosing cholangitis, cirrhosis, inflammatory bowel disease and smoking status

Table S7. Characteristics of case-control studies conducted in the US on the association between HBV or HCV infection and intrahepatic and/or extrahepatic bile duct cancer (continued)

Study[a]	Study population	Matching variables	Characteristics of participants, age in years (SD)	Detection method	Cancer	Cases		Controls		OR[b] (95% CI)	Adjustment variables & remarks
						n/N	Pos %	n/N	Pos %		
Shaib 2007[250]	Cases: cholangiocarcinoma patients referred to the M.D. Anderson Cancer Center between 1992 and 2002 Controls: randomly selected from an existing database of healthy individuals (genetically unrelated family members, spouses and friends of patients who had cancer other than gastrointestinal cancer) interviewed between 1999–2004 at M.D. Anderson	Frequ-ency matched by gender, ethnicity and age (±5 yr)	ICC cases: 59.8 (11.4) ECC cases: 61.1 (9.8) Controls: 58.1 (11.4)	**HBsAg**	ICC	1/83	1.2	1/236	0.4	2.9 (1.97–4.46)	Race, age, gender, HCV & HBV markers, heavy drinking Lower bound CI was imputed
					ECC	4/163	2.5	1/236	0.4	1.4 (0.01–56.5)	
				Anti-**HCV**	ICC	5/83	6.0	2/236	0.8	7.9 (1.3–84.5)	
					ECC	6/163	3.7	2/236	0.8	2.8 (0.3–35.1)	
Shaib 2005[263]	SEER-Medicare database Cases: persons diagnosed no earlier than 1993 and who had two years of Medicare data before the date of diagnosis and up to one year after ICC diagnosis or until death Controls: 5% random sample of Medicare-enrolled beneficiaries with no cancer of any type residing in the geographic regions of SEER registries Period: 1993–1999	Years of search for risk factors	>65 Cases: 78.7 (6.4) Controls: 76.5 (6.9)	ICD-9 codes for **HBV**	ICC	1/625	0.2	181/90,834	0.2	0.8 (0.1–5.9)	Age, sex, race, geographic region & Medicare/Medicaid dual enrollment
				ICD-9 codes for **HCV**	ECC	5/625	0.8	161/90,834	0.2	5.2 (2.1–12.8)	

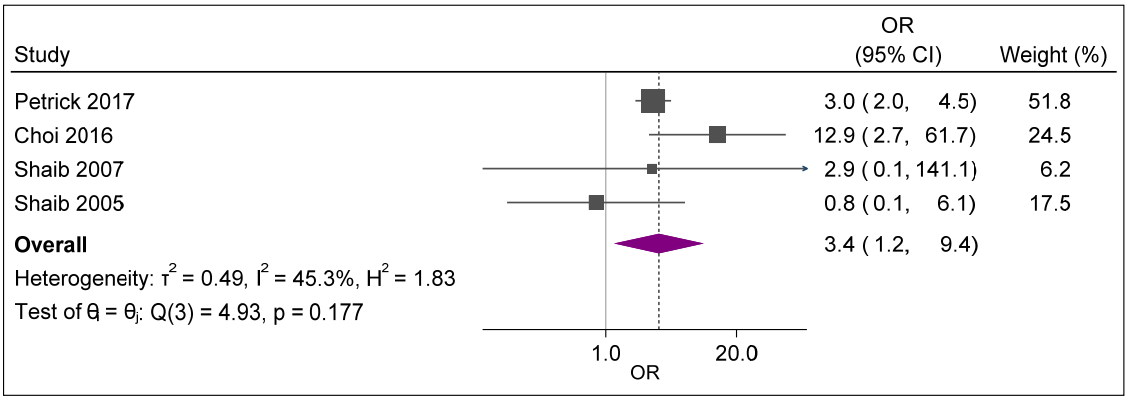
CI = confidence interval, ECC = extrahepatic cholangiocarcinoma, HBV = Hepatitis B virus, HBsAg = hepatitis B surface antigen, HCV = hepatitis C virus, ICC = intrahepatic cholangiocarcinoma, Pos = positive, SD = standard deviation, SEER = Surveillance, Epidemiology, and End Results Program

a. Inclusion criteria: hepatitis infection confirmed by serology (HBsAg, anti-HCV [with or without RIBA confirmation], HCV RNA), 10 or more ICC or ECC cases, study population from the US

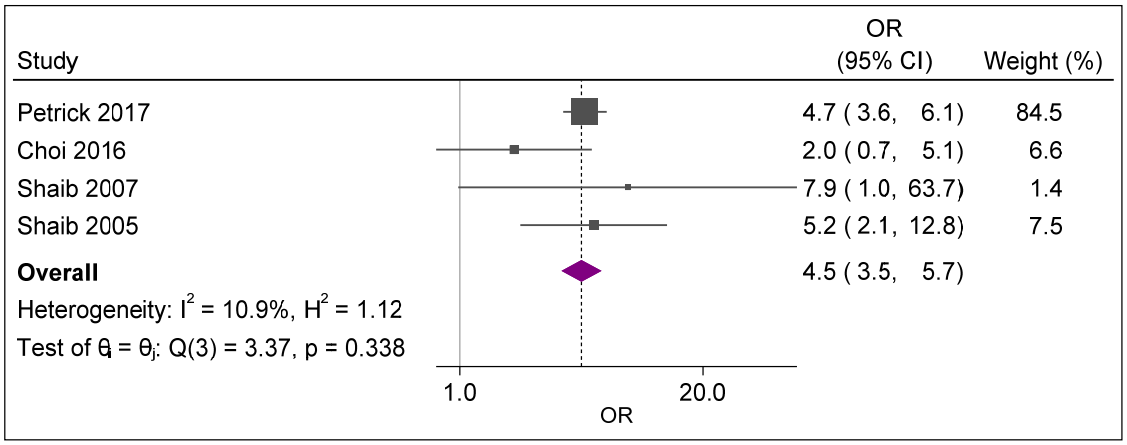
b. After the normalizing transformation is performed, the CIs listed in the table may not match those in the forest plot.

Fig. S2. Pooled ORs for the association between each (1) HBV and (2) HCV and intrahepatic bile duct cancer

(1) Hepatitis B virus

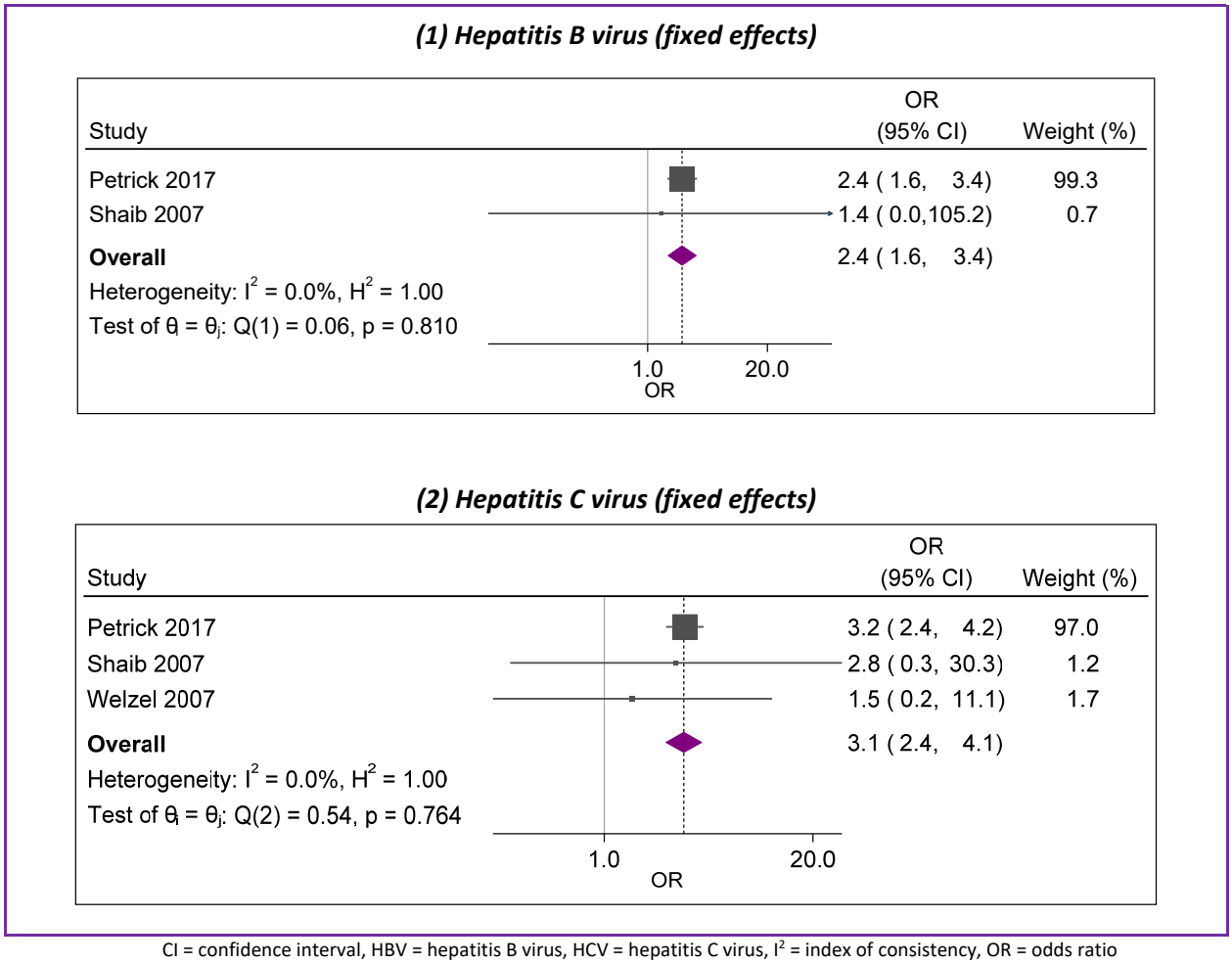


(2) Hepatitis C virus (fixed effects)



CI = confidence interval, HBV = hepatitis B virus, HCV = hepatitis C virus, I^2 = index of consistency, OR = odds ratio

Fig. S3. Pooled ORs for the association between each (1) HBV and (2) HCV and extrahepatic bile duct cancer



CI = confidence interval, HBV = hepatitis B virus, HCV = hepatitis C virus, I^2 = index of consistency, OR = odds ratio

HELICOBACTER PYLORI

H. pylori is estimated to infect about 50% of the world's population, but the prevalence varies globally, likely reflecting socio-demographic and economic conditions of the regions.[35] Although infection is mostly acquired during childhood, *H. pylori* prevalence increases with age.[275] A decrease in the overall prevalence of *H. pylori* infection has been observed in recent years, with successive generations presenting lower prevalence.[35] The one cycle of the NHANES assessed *H. pylori* serostatus collected data from participants aged ≥3 years, via ELISA from 1999–2000. Equivocal results (representing <2% of results) were categorized as positive (**Table S8**).

Table S8. Estimated *Helicobacter pylori* prevalence in the US and PAFs for NCGC

Sex-age group[a]	NHANES estimates from the 1999–2000 cycle									PAF estimates for NCGC for 2017 (95% CI)
	Sample				Weighted			Imputed + Weighted		
	Pos no.	Pos %	RSE %	Missing[b] %	Pos %	RSE %	Missing %	Pos %	RSE %	
Males										
10–14	134	23.5	7.6	9.4	15.0	16.9	13.3	14.6	16.5	Not included
15–19	211	32.4	5.7	8.2	16.2	9.9	11.2	16.4	10.8	Not included
20–24	42	27.1	13.2	6.6	19.9	14.9	3.7	20.4	14.5	63.3 (47.2–74.5)
25–29	54	35.1	11.0	4.9	28.8	12.1	4.3	28.9	11.8	65.9 (52.4–75.5)
30–34	67	43.5	9.2	5.5	31.1	10.2	5.1	31.1	9.7	70.6 (56.8–80.0)
35–39	71	43.0	9.0	5.7	27.1	12.6	4.1	27.2	11.9	77.3 (66.1–84.7)
40–44	98	52.1	7.0	6.0	35.6	14.6	5.8	34.5	14.7	78.5 (68.4–85.4)
45–49	69	51.9	8.4	5.7	29.3	19.0	4.6	29.6	19.1	76.2 (64.7–83.9)
50–54	71	52.2	8.2	6.9	37.7	7.4	8.0	37.8	7.9	80.2 (69.3–87.3)
55–59	54	50.0	9.6	5.3	37.8	15.2	1.7	38.4	15.3	77.7 (64.2–86.1)
60–64	111	56.6	6.3	3.9	38.2	17.7	4.0	38.8	17.7	81.6 (72.9–87.6)
65–69	82	50.9	7.7	8.0	35.3	11.4	6.9	36.7	11.2	81.9 (71.4–88.5)
70–74	73	50.3	8.2	7.6	37.1	15.5	10.1	36.8	14.8	82.0 (70.9–88.9)
75–79	57	55.9	8.8	8.1	43.4	13.7	5.9	44.1	12.9	81.2 (71.6–87.6)
80–84	48	51.1	10.1	9.6	50.7	10.9	11.3	51.0	11.7	81.2 (70.7–88.0)
≥85	--	--	--	--	--	--	--	--	--	83.8 (74.9–89.6)
Overall	**1242**	**39.9**	**2.2**	**7.3**	**29.2**	**4.0**	**6.4**	**29.2**	**4.0**	**80.0 (69.1–87.0)**
Females										
10–14	103	17.9	8.9	12.2	9.2	19.7	14.4	9.3	18.9	Not included
15–19	172	27.8	6.5	7.4	16.5	14.4	9.8	17.1	14.5	Not included
20–24	79	33.1	9.2	7.0	20.9	18.3	7.7	20.8	18.2	52.2 (34.2–65.3)
25–29	59	28.4	11.0	8.4	20.0	15.1	8.7	19.8	14.8	66.8 (52.1–77.0)
30–34	73	34.0	9.5	4.9	27.4	13.7	4.3	27.3	13.1	71.0 (55.7–81.0)
35–39	76	40.9	8.8	6.1	26.5	14.6	3.9	27.1	14.8	70.0 (55.8–79.6)
40–44	86	47.5	7.8	5.2	26.8	14.2	6.2	27.2	13.8	76.2 (64.4–84.1)
45–49	74	43.8	8.7	8.2	29.9	8.7	5.7	30.9	8.7	76.1 (63.7–84.3)
50–54	79	48.2	8.1	5.8	37.3	10.9	5.6	38.2	11.1	76.2 (64.1–84.1)
55–59	64	55.7	8.3	2.5	38.8	14.1	2.7	38.8	14.0	78.4 (68.4–85.2)
60–64	91	50.3	7.4	9.5	38.6	15.3	7.5	39.4	14.6	81.8 (72.4–88.0)
65–69	94	59.5	6.6	10.2	46.1	10.4	9.1	45.6	11.0	82.0 (72.0–88.4)
70–74	82	57.3	7.2	5.3	49.9	11.0	5.4	49.6	11.4	82.3 (72.2–88.7)
75–80	53	54.1	9.3	5.8	48.2	15.3	3.9	48.3	14.9	84.3 (75.9–89.8)
80–84	39	40.2	12.4	7.6	30.7	10.9	8.5	32.9	12.5	85.4 (77.4–90.5)
≥85	--	--	--	--	--	--	--	--	--	85.0 (76.1–90.6)
Overall	**1224**	**36.6**	**2.3**	**7.9**	**28.2**	**4.6**	**7.0**	**28.3**	**4.7**	**79.9 (69.5–86.7)**

CI = confidence interval, NHANES = National Health and Nutrition Examination Survey, NCGC = non-cardia gastric cancer, PAF = population attributable fraction, Pos = positive, RSE = relative standard error, US = United States

-- *H. pylori* prevalence among those ≥85 (in 1999–2000) was not calculated because it was not required after applying a latency period.

a. We did not consider *H. pylori* infection prevalence among those aged 3 to 9 (26% did not have a test result).

b. Missing refers to individuals who attended the interview and medical examination but do not have a test result for *H. pylori* infection.

Gastric cancer (non-cardia)

This cancer is often classified according to its physical location within the stomach: tumors located in the upper region of the stomach, specifically within 1 to 2 cm proximal and 2 cm distal to the esophagogastric mucosal junction, are identified as cardia cancers; cancers located in the fundus, body, pyloric antrum or pylorus regions are identified as non-cardia.[276] The latter are the most frequent, accounting for 61% of the cases diagnosed in the US in 2012 (males: 51.8%, females: 75.5%).[277] *H. pylori* infection is known to increase the risk of non-cardia gastric cancer with a reported pooled estimate of 2.81 (CI: 2.14–3.68) considering case-control studies and case-control studies nested within prospective cohorts.[278] The association between *H. pylori* infection and gastric cardia adenocarcinomas remains conflicting. Studies from low gastric cancer risk settings, namely Europe, the US and Australia, generally report null or inverse associations (pooled RR = 0.78, CI: 0.63–0.97), while statistically significant associations have been observed in high-risk settings, namely China, Japan and Korea (pooled RR = 1.98, CI: 1.38–2.83).[278] A recent case-cohort study from China, an area of high *H. pylori* infection endemicity, obtained a statistically significant association (hazard ratio = 3.06, CI: 1.54–6.10).[276] These differences and null associations observed may be explained by the coexistence of two distinct types of cardia gastric cancer.[130] One arises from non-atrophic gastric mucosa, associated with acid/bile-induced damage to the distal esophagus, resembling esophageal adenocarcinoma[279] and is likely to have a higher relative frequency in settings with low overall gastric cancer risk. The other is associated with *H. pylori* induced atrophic gastritis,[279] which is etiologically similar to non-cardia tumors and more frequent in populations with a high frequency of gastric cancer. It is possible that *H. pylori* infection may be associated with a small fraction of cardia gastric cancer, however it is difficult to determine the origin of these cancers to obtain an accurate estimate.

In retrospective studies, individuals with gastric cancer may test negative following the clearance of infection associated with atrophic gastritis, thus underestimating the prevalence of *H. pylori* infection among cases.[277] As such, only cohort studies or case-control studies nested within prospective cohorts were considered to estimate the association between *H. pylori* infection and non-cardia gastric cancer.

The finding that immunoblot is more sensitive than ELISA/enzyme immunoassay (EIA)[278] in detecting *H. pylori* necessitated a correction for this potential error.[46] The sensitivity and specificity were extracted, and pooled from two studies that compared ELISA to immunoblot head-to-head.[106,107] A derivation of a formula used to correct measurement error (91% sensitivity and 95% specificity) in the ORs was applied to the five nested case-controls that used EIA or ELISA (**Table S9**).[105] The corrected and immunoblot studies (**Table S10**) were pooled with fixed effects due to a lack of heterogeneity (**Fig. S4**).

Table S9. Characteristics of studies on the association between *H. pylori* infection detected using ELISA or EIA and non-cardia gastric cancer

Study[a]	Study population	Follow-up years, mean/ median	Matching variables	Characteristics of participants, ages in years	Cases		Controls		Unadjusted OR (95% CI)	Corrected OR[b,c] (95% CI)
					n/N	Pos %	n/N	Pos %		
Persson 2011[252]	Swedish cohorts (Swedish Institute for Infectious Disease Control Biobank and Malmö Microbiology Biobank) Recruited: 1968–2001 Diagnosed: 1968–2006	16.5	Sex, age, sera collection year, biobank	Mean (SD; range) age at sera collection: 30.8 (6.1; 16–40) for cases; 30.9 (6.0; 16–40) for controls Mean (SD; range) age at diagnosis: 47.3 (9.4; 25–68)	35/41	85.4	30/81	37.0	9.9 (3.7–26.3)	21.5 (6.1–75.8)
Hansen 2007[130]	Norwegian cohort (Janus Serum Bank Cohort) Recruited: 1972–1986 Diagnosed: 1972–1992	11.9	Sex, age, cohort, sera collection date and study source	Males: 91 cases; 267 controls Median (range) age at sera collection: 45.6 (23.6–63.4) Median (range) age at diagnosis: 55.8 (34.3–68.2)	116/129	89.9	247/376	65.7	4.7 (2.5–8.6)	26.6 (6.5–109.1)
Knekt 2006[131]	Finnish cohort (Finnish Mobile Clinic Health Examination cohort) Recruited: 1968–1972 Diagnosed: 1968–1991	Up to 24	Sex, age, municipality	Males: 120 cases; 231 controls Mean age (SD) at baseline: 68 (14) for cases	176/193	91.2	292/372	78.5	2.8 (1.6–4.9)	66.2 (4.1–1078.6)
Nomura 2002[132]	US cohort of men of Japanese ancestry Recruited: 1967–1977 Diagnosed: 1967–1996	12.7	Age, sera collection date	All men Mean (range) age at diagnosis: 72.5 (50.2–90.3)	231/261	88.5	193/261	73.9	2.7 (1.7–4.3)	7.9 (3.7–16.9)
Parsonnet 1993[253]	US cohort of adult subscribers to the Kaiser Permanente Medical Care Program Recruited: 1964–1969 Diagnosed: 1964–1989	15	Sex, age group, race, sera collection date and study site	Median age at sera collection: 53.6	84/98	85.7	61/98	62.2	3.6 (1.8-7.3)	7.4 (3.1–19.6)

CI = confidence interval, EIA = enzyme immunosorbent assay, ELISA = enzyme-linked immunosorbent assay, OR = odds ratio, Pos = positive, SD = standard deviation, US = United States

a. Inclusion criteria: prospective serology collection (~10 years in advance of diagnosis), ELISA or EIA, 10 or more non-cardia gastric cancer cases, North American, European or Australian and New Zealand study populations, data required to correct sensitivity and specificity.

b. Corrected to 91% sensitivity and 95% specificity. ORs were calculated based on the condition maximum likelihood estimates, and CIs were based on Fisher exact tests.

c. After the normalizing transformation is performed, CIs listed in the table may not match those in the forest plot.

Table S10. Characteristics of studies on the association between *H. pylori* infection detected using immunoblot and non-cardia gastric cancer

Study[a]	Study population	Follow-up yrs, mean/ median	Matching variables	Characteristics of participants, ages in years	Cases		Controls		Adjusted OR[b] (95% CI)	Adjustment variables
					n/N	Pos %	n/N	Pos %		
Gonzalez 2012[106]	10 European countries in the EPIC cohort Recruited: 1992–1998 Diagnosed: 2000–2004	10.7	Sex, age group, study center, date of blood collection	Age range at baseline: 40-65	82/88	93.2	199/338	58.9	21.4 (7.1–64.4)	Smoking status, school level, red and processed meat intake, fruit & vegetables consumption
Mitchell 2008[107]	Australian cohort (Melbourne Collaborative Cohort Study) Recruited: 1990–1994 Diagnosed: 1990–2002	11.6	Sex, age, birth country, sera collection date	Males: 21 cases; 84 controls Median (range) age at baseline: 62 (42–69)	32/34	94.1	85/134	63.4	10.6 (2.4–47.4)	None
Simán 2007[133]	Swedish cohort (Malmö Preventive Medicine) Recruited: 1974–1992 Diagnosed: –2000	Ranged from 9.2–12.6	Sex, age, sera collection date	Males: 54 cases Mean (range) age at baseline: 50.7 (34.0–60.9)	65/67	97.0	147/250	58.8	17.8 (4.2–74.8)	Occupation, tobacco consumption

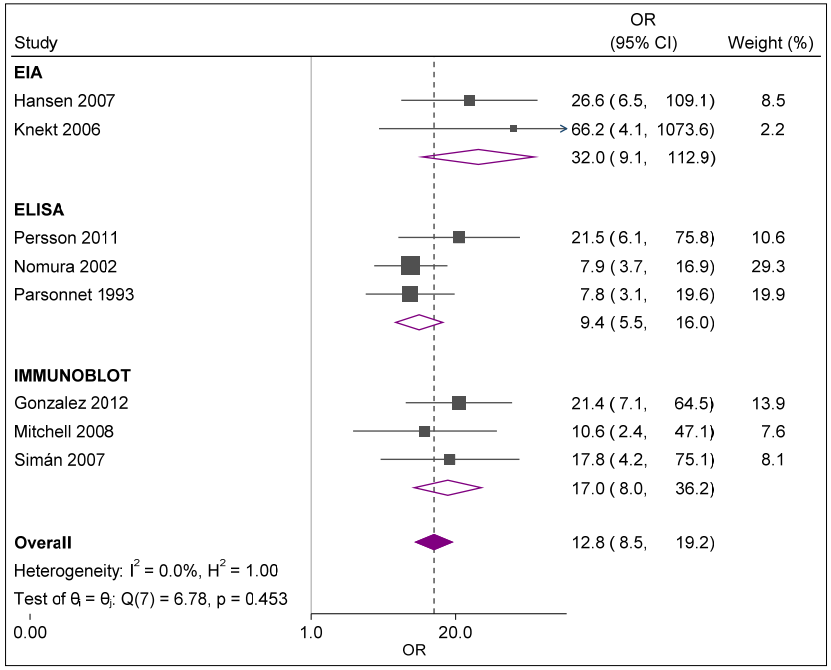
CI = confidence interval, EPIC = European Prospective Investigation into Cancer and Nutrition, OR = odds ratio, Pos = positive

a. Inclusion criteria: prospective serology collection (~10 years in advance of diagnosis), immunoblot detection, 10 or more non-cardia gastric cancer cases, North American, European or Australia and New Zealand study populations.

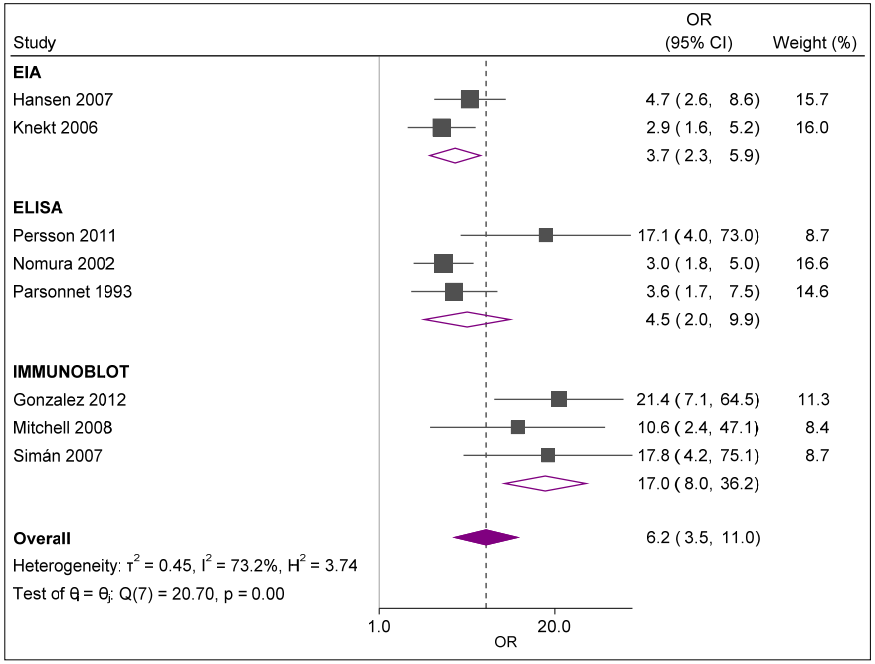
b. After the normalizing transformation is performed, CIs listed in the table may not match those in the forest plot.

Fig. S4. Pooled corrected (1) and uncorrected (2) ORs for the association between H. pylori and non-cardia gastric cancer

(1) Corrected ORs: used in analysis (fixed effects)



(2) Uncorrected ORs: for comparison purposes



CI = confidence interval, EIA = enzyme immunoassay, ELISA = enzyme-linked immunosorbent assay, I^2 = index of consistency, OR = odds ratio

Gastric mucosa-associated lymphoid tissue (MALT) lymphoma and DLBCL

MALT lymphoma, a type of non-Hodgkin lymphoma (NHL), is most often diagnosed in the stomach, but can also be found in the lungs, thyroid, skin or soft tissues.[134] A systematic review of published series found that *H. pylori* infection is present in nearly 90% of patients with gastric MALT lymphoma.[135] According to current guidelines, antibiotic therapy against *H.* pylori infection is the first-line of treatment in patients with gastric MALT regardless of stage of disease and prognosis factors.[280,281] In fact, *H. pylori* eradication confers a ~74% remission rate of MALT in Western populations.[135] Even among patients with *H. pylori*-negative gastric MALT, complete remission following eradication therapy is nearly 30%.[282] The data on the association between *H. pylori* and gastric MALT were very limited, we identified only one cohort study examining the relationship between *H. pylori* and gastric NHL. This study, conducted by Parsonnet et al., combined data from two cohort studies conducted in Norway and the US,[122] and reported a measure of association (OR = 6.3, CI: 2.0–19.9) for NHL of gastric location. This study included 33 cases matched to four controls by cohort, sex, age and sera collection date.[122] Of the 33 cases, just three cases were gastric MALTs, one case was lymphocytic lymphoma, and the remaining 29 cases were DLBCLs. We opted to utilize the OR for the US cohort (7.9, CI: 1.6–38.1) which included 20 gastric NHL cases and apply it to gastric MALT and DLBCL incidence.

Esophageal adenocarcinoma

Esophageal cancer presents with two major histological types: squamous cell carcinoma (morphology codes 8140-8576) that most often arises in the middle third of the esophagus, followed by the lower and the upper third, and adenocarcinoma (morphology codes 8050-8083) that usually develops in the lower third.[283] In the US, esophageal adenocarcinoma accounted for 55% of esophageal cancer cases diagnosed between 2001 and 2015.[284] *H. pylori* infection is inversely associated with the occurrence of esophageal adenocarcinoma, regardless of other environmental and genetic exposures,[283,285] and the decline in the prevalence of *H. pylori* infection may have contributed to an increase in esophageal adenocarcinoma incidence. The effect of *H. pylori* infection has been evaluated by several meta-analyses reporting results for both esophageal squamous cell carcinoma and adenocarcinoma.[221-226] All reported similar results, showing no association between *H. pylori* and esophageal squamous cell carcinoma, while for adenocarcinoma a protective effect of *H. pylori* infection was found (OR ≈ 0.5).

The mechanism through which *H. pylori* infection reduces the risk of esophageal adenocarcinoma is not yet clear. Studies have suggested *H. pylori* infection may decrease gastric cancer secretion by acting on parietal cells via bacterial products and cytokines or through mucosal atrophy resulting from chronic inflammation. Consequently, there may be less reflux esophagitis, Barrett's esophagus, and development of esophageal adenocarcinoma.[254,286] However, the association between the absence of *H. pylori* infection and increased gastroesophageal reflux,[287] and whether infection interacts directly with host epithelial cells and/or affects the microbial composition of the esophagus remain unclear.[288] Nevertheless, previous studies have suggested that the association between *H. pylori* infection and esophageal adenocarcinoma may be independent of CagA status and atrophy of the stomach.[254,257,289]

Our search produced six meta-analyses that reported results for the association between *H. pylori* infection and esophageal adenocarcinoma, all reported a protective effect.[221-226] Ten individual studies were conducted in the US.[254-257,290-295] Studies that did not provide estimates for esophageal adenocarcinoma specifically (i.e., considered esophageal and gastric cardia adenocarcinoma,[291] Barrett's esophagus complicated by dysplasia or adenocarcinoma)[290,295] and/or considered controls with gastrointestinal symptoms or undergoing endoscopy for reasons other than screening (i.e., patients undergoing endoscopy due to achalasia, familial adenomatous polyposis, chronic diarrhea, lower abdominal pain, Hemoccult-positive stools, unexplained nausea and vomiting, and unexplained chest pain;[290,295] patients with benign disease and symptoms suggestive of foregut disease;[294] patients undergoing esophagogastroduodenoscopy for classic symptoms of GERD with or without complaints of dysphagia, nocturnal cough, chest pain, nausea, vomiting, or signs of acute or chronic gastrointestinal bleeding;[292] patients with intestinal metaplasia)[293] were excluded.[290,295] Four studies met the inclusion criteria (**Table S11**) and were pooled with fixed effects due to a lack of heterogeneity (**Fig. S5**).

Table S11. Characteristics of studies on the association between *H. pylori* infection and esophageal adenocarcinoma conducted in the US

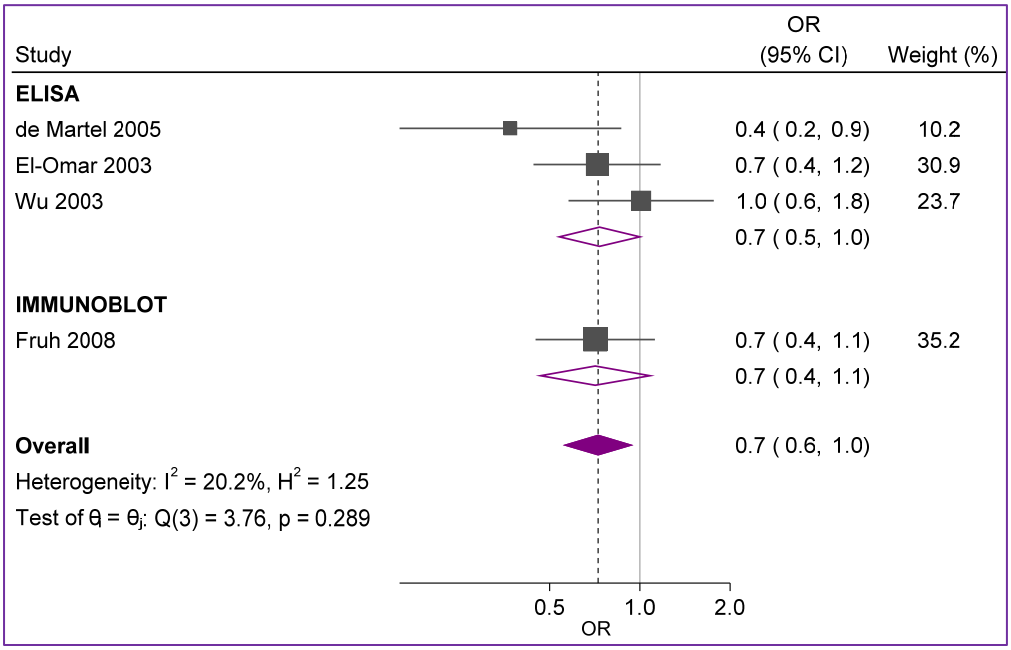
Study[a]	Study population	Matching variables	Characteristics of participants, ages in years	Assess-ment of *H. pylori* infection	Cases		Controls		Adjusted OR (95% CI)[b]	Adjustment variables
					n/N	Pos %	n/N	Pos %		
Fruh 2008[257]	Case-control study Cases: histologically confirmed esophageal adenocarcinoma patients at the Massachusetts General Hospital Controls: selected from healthy GERD-free, non-blood-related family members and friends of other cancer/surgical patients Diagnosed/Recruited: not specified	Sex, age	Males: 88 cases; 88 controls Mean age (SD): 64 (8) for cases; 63 (8) for controls	Serum (Helicoblot)	36/100	36.0	43/101	42.6	0.71 (0.4–1.0)	Adult body mass index, smoking status, age, sex
de Martel 2005[254]	Nested case-control study (Kaiser Permanente Medical Care Program) Cases: esophageal adenocarcinoma patients were identified in the cohort and were confirmed by information in the SEER database Controls: randomly selected from the cohort Recruited: 1964–1969 Diagnosed: 1964–2000	Sex, age, race, date & site of sera collection	Males: 41 cases; 121 controls Mean age (SD): 47.9 (10.0) for cases; 47.7 (9.6) for controls	Serum IgG (ELISA)	19/51	37.3	74/150	51.0	0.37 (0.16-0.88)	Body mass index, cigarette smoking, education
Wu 2003[256]	Case-control study Cases: esophageal adenocarcinoma patients from the Los Angeles County Cancer Surveillance Program (population-based cancer registry) Controls: selected from the neighborhood of residence of the case patient Diagnosed/Recruited: 1992–1997	Sex, age group, race	Males: 73 cases; 261 controls	Serum IgG (ELISA)	49/80	61.2	230/356	64.6	1.01 (0.58-1.77)	Sex, age, education, birthplace, ethnic group, smoking status, body mass index
El-Omar 2003[255]	Case-control study Cases: esophageal adenocarcinoma patients from New Jersey and western Washington Controls: population-based controls selected by random-digit dialing and from Health Care Financing Administration files Diagnosed/Recruited: 1993–1995	Sex, age group, study centre	Males: 93 cases; 178 controls Median age: 65 for cases; 66 for controls	Serum IgG (ELISA)	35/108	32.4	84/210	40.0	0.72 (0.44-1.17)	None

CI = confidence interval, ELISA = enzyme-linked immunosorbent assay, GERD = gastroesophageal reflux disease, Pos = positive, OR = odds ratio, SD = standard deviation, SEER = Surveillance, Epidemiology, and End Results, US = United States

a. Inclusion criteria: cohort, nested case-control or case-control studies with *H. pylori* infection confirmed by serology (ELISA, enzyme immunoassay [EIA] or immunoblot), 10 or more esophageal adenocarcinoma cases, controls without gastrointestinal symptoms and not undergoing endoscopy for purposes other than screening, study population from the US.

b. After the normalizing transformation is performed, the CIs listed in the table may not match those in the forest plot.

Fig. S5. Forest plot of the association between H. pylori infection and esophageal adenocarcinoma (fixed effects)[a,b]



CI = confidence interval, ELISA = enzyme-linked immunosorbent assay, I^2 = index of consistency

a. The study by de Martel and colleagues published in 2005 is a nested case-control where *H. pylori* sera collection occurred prior to adenocarcinoma diagnosis, the remaining studies are case-controls.

b. Pooling the unadjusted ORs from the four studies resulted in a pooled OR of 0.75 (CI: 0.57–0.98).

EPSTEIN-BARR VIRUS (EBV)

Carcinogenicity is demonstrated by the detection of EBV viral genome within the tumor cells (i.e., where the EBV genome is translated and transcribed).[139] To detect EBV within cancer tissues, EBV-encoded RNA *in situ* hybridization (EBER ISH) is viewed as the gold standard assay;[1,120] for Hodgkin lymphoma (HL), latent membrane protein 1 (LMP-1) is comparable to EBER.[120]

Burkitt lymphoma (BL)

BL in children (aged 0–9 years)

We identified seven studies conducted in the US and Europe (**Table S12**). The pooled prevalence of EBV was 15.5% (CI: 8.1–23.0%) (**Fig. S6**).

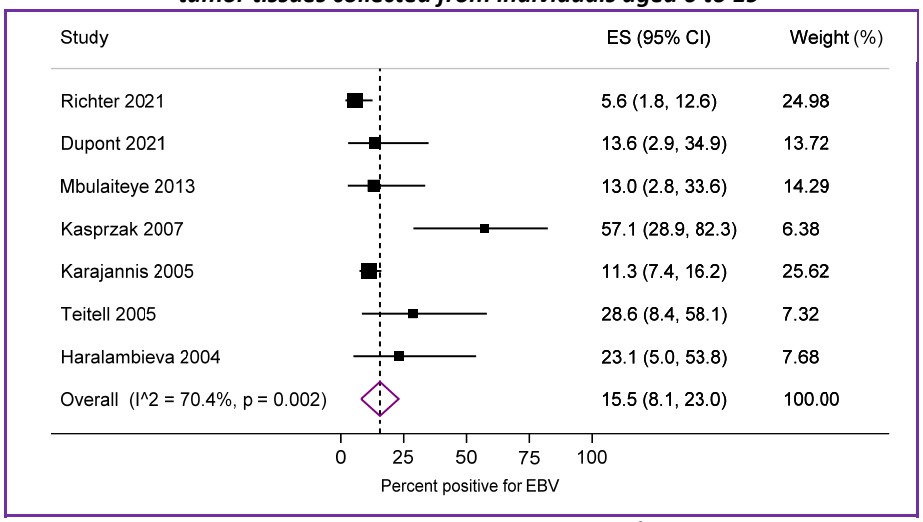
Table S12. Characteristics of studies on EBV prevalence in Burkitt lymphomas from individuals aged 0 to 19

Study[a]	Region(s)	Source of cases	Diagnosis dates	Male %	Age range (years)	Tested n/N	Pos %
Richter 2021[296]	Germany	Hematopathology Section and Lymph Node Registry of the University Hospital Schleswig-Holstein, Campus Kiel	2001–2013	86.8	≤18	5/89	5.6
Dupont 2021[297]	Denmark	Danish Registry of Pathology	1980–2018	81.8	3–19	3/22	13.6
Mbulaiteye 2013[140]	Los Angeles County, Hawaii & Iowa	Residual tissue repositories (population-based) and diagnostic referral centers	1979–2009	91.3	0–19	3/23	13.0
Kasprzak 2007[298]	Poland	Department of Haematology and Paediatric Oncology	1999–2003	92.9	3–16	8/14	57.1
Karajannis 2005[299]	Austria, Germany & Switzerland	NHL-BFM (Berlin-Frankfurt-Munster) data center	1990–1998	79.7	1–18	25/222	11.3
Teitell 2005[300]	France & United Kingdom	Institut Gustave Roussy & Children's Hospital	NS	85.7	2–16	4/14	28.6
Haralambieva 2004[301]	the Netherlands	Pathology departments & Dutch Childhood Oncology Group	NS	NS	5–13	3/13	23.1

EBV = Epstein-Barr virus, EBER ISH = Epstein-Barr encoding region *in situ* hybridization, NS = Not specified, Pos = positive

a. Inclusion criteria: tissue specimen tested for EBV, EBER ISH detection, European or North American cases, and eight or more participants

Fig. S6. Forest plot of EBV prevalence (%) in Burkitt lymphoma tumor tissues collected from individuals aged 0 to 19



CI = confidence interval, EBV = Epstein-Barr virus, ES = effect size, I^2 = index of consistency

BL in adults (aged ≥20 years)

We identified two studies conducted in the US that utilized EBER ISH; they were one study by Mbulaiteye and colleagues (2014) of 40 human immunodeficiency virus (HIV)-negative or unknown HIV status cases (11 HIV+ cases excluded by us) diagnosed from 1979–2009 using Surveillance, Epidemiology, and End Results (SEER) data collected from Los Angeles County, Hawaii and Iowa, where 27.5% tested EBV positive (including the 11 people living with HIV [PLWH], 35.3% tested positive).[140] Another study by Naeini and colleagues (2016), tested 27 BL cases of unknown HIV status sent to pathology services in California, and reported that 10 (37.0%) tested positive.[267]

Pooling five studies (four conducted in Europe and one in the US)[140,302-305] reporting on EBV prevalence in 118 BLs among PLWH, provided prevalence of 50.1% (CI: 34.6–65.6; data not shown). Considering individuals aged 20–59 (since the estimated proportion of BLs occurring among PLWH aged ≥60 years was only 2.0% over 1980–2007, we did not consider this age group) an estimated 21.5% of BLs from the most recent period available (2001–2007) were diagnosed among PLWH in the US.[237] Weighting the pooled prevalence by HIV status provided EBV prevalence of 35.1%, which is near identical pooled EBV prevalence reported by Mbulaiteye and colleagues that included general and PLWH cases. For this reason, we instead opted to use age-group specific EBV prevalence from the Mbulaiteye study which included some HIV+ cases: 55% (aged 20–34), 33% (aged 35–59), 25% (aged ≥60).[140]

Hodgkin lymphoma (HL)

HL in children (aged 0–19 years)

Pooling six studies that provided EBV prevalence for younger versus older children (**Table S13**), resulted in EBV prevalence of 62.2% for children aged 0–9 and 22.3% for those aged 10–19 (**Fig. S7**).

HL in adults (aged ≥20 years)

Pooling four studies reporting on EBV prevalence in two adult age groups, provided a pooled prevalence of 20.5% in adults aged 15–44 years old and 42.5% in adults aged ≥45 years old (**Fig. S8**). Pooling six studies (two[306,307] conducted in the US and four[308-311] in Europe) reporting on

EBV prevalence in 282 HL cases diagnosed among PLWH, resulted in prevalence of 92.9%. Using data from the 14 SEER cancer registries (2000–2010), Shiels and colleagues estimated the proportion of HLs among PLWH by sex and age group.[236] We utilized the proportion of HL cases estimated to be among PLWH by 10-year age groups from age 20 to 69 to partition HL cancer incidence; these proportions were 1.5% (age 20–29), 5.4% (age 30–39), 9.3% (age 40–49), 7.3% (age 50–59), and 1.9% (age 60–69) and applied to males HL incidence counts only.[236]

Table S13. Characteristics of studies conducted in the US, Canada, *or Europe (for HIV+ only)*, reporting on EBV prevalence in Hodgkin lymphoma

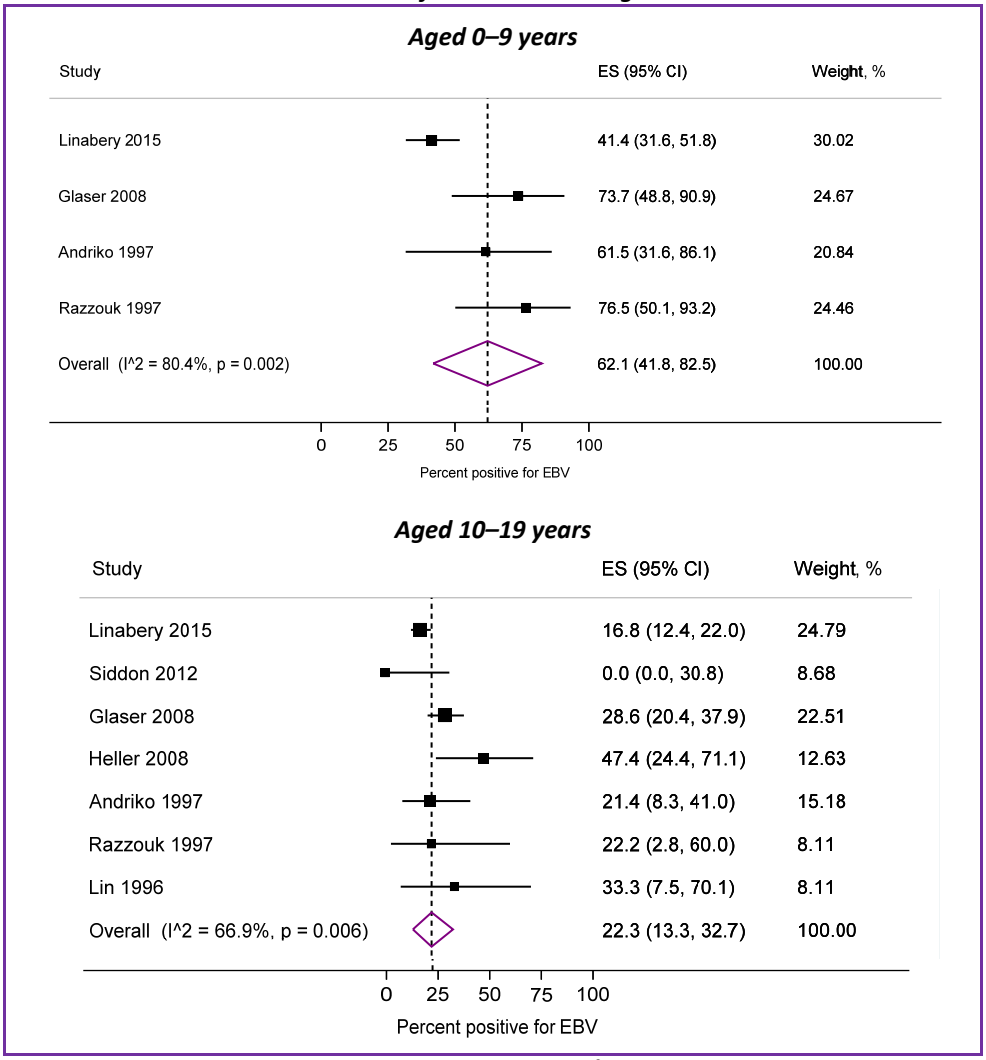
Study[a]	Region(s)	Source of cases	Diagnosis dates	Male %	Detection method(s)	HIV status	Age range (years)	Tested n/N	Pos %
Linabery 2015[312]	United States, Puerto Rico & Canada	Children's Oncology Group	1989–2003	NS	EBER ISH	Unknown	0–9	41/99	41.4
							10–14	43/256	16.8
Siddon 2012[313]	Connecticut	Yale-New Haven Hospital	NS	50.0	EBER ISH	HIV-	10–19	0/10	0.0
Glaser 2008[314]	California	California Cancer Registry and non-White Los Angeles County residents	1988–1997	NS	EBER ISH, LMP-1	HIV-	0–9	14/19	73.7
							10–19	32/112	28.6
							20–49	136/650	20.9
							≥50	122/251	48.6
Heller 2008[315]	New York	Memorial Sloan-Kettering Cancer Center	NS	45.5	EBER ISH	HIV-	10–19	9/19	47.4
Chang 2004[147]	Massachusetts & Connecticut	Population-based case-control study	1997–2001	57.0	EBER ISH, LMP-1	HIV-	15–44	55/291	18.9
							≥45	41/108	38.0
Vasef 2004[148]	Iowa	Pathology department	NS	58.6	EBER ISH, LMP-1	Unknown	15–44	6/24	25.0
Andriko 1997[316]	Washington, D.C.	Lymphatic Pathology Registry, Armed Force Institute of Pathology	1984–1996	90.9	LMP-1	Unknown	0–9	8/13	61.5
							10–19	6/28	21.4
Razzouk 1997[317]	Tennessee	St. Jude Children's Research Hospital	NS	42.3	EBER ISH	Unknown	0–9	13/17	76.5
							10–19	2/9	22.2
Elenitoba-Johnson 1996[149]	Rhode Island	Pathology departments	NS	42.9	LMP-1	Unknown	15–44	6/18	33.3
							≥45	3/10	30.0
Lin 1996[318]	Maryland	Clinical Center of National Institutes of Health	1971–1992	NS	EBER ISH	Unknown	10–19	3/9	33.3
Studies conducted among adults living with HIV									
Besson 2015[308]	France	22 centres: French Cohort of HIV-related lymphomas—French National Agency for Research on AIDS and Viral Hepatitis ANRS-CO16 Lymphovir cohort	2008–2014	86.8	EBER-1 ISH, LMP-1	HIV+	38–48	39/42	92.9
Hentrich 2012[309]	Austria & Germany	42 institutions in Austria & Germany	2004–2010	92.6	LMP (81%), EBER ISH (4%), PCR (3%), LMP & EBER (4%), method NS (9%)	HIV+	27–70	95/103	92.2
Glaser 2003[306]	California	California Cancer Registry non-White Los Angeles County residents	1988–1998	100.0	EBER ISH, LMP-1	HIV+	NS	53/59	89.8
Thompson 2004[307]	D.C.	AIDS Registry of the Armed Forces Institute of Pathology[b]	1984–2000	97.8	LMP	HIV+	21–75	32/33	97.0
Carbone 1999[310]	Italy	NS	NS	NS	EBER ISH	HIV+	NS	25/27	92.6
Tirelli 1995[311]	Italy	Division of Pathology at the Centro di Riferimento Oncologico	NS	NS	EBER-1 & EBER-2 ISH, Southern blotting	HIV+	NS	14/18	77.8

AIDS = acquired immunodeficiency syndrome, D.C. = District of Columbia, EBV = Epstein-Barr virus, EBER ISH = Epstein-Barr encoding region *in situ* hybridization, HIV = human immunodeficiency virus, LMP = latent membrane protein, NS = Not specified, PCR = polymerase chain reaction, Pos = positive, US = United States

a. Inclusion criteria: tissue specimen tested for EBV, EBER ISH detection, North American cases (for HIV+ cases from Europe were also eligible), and ≥8 cases (children) or ≥10 (adults), EBV prevalence reported by age-group.

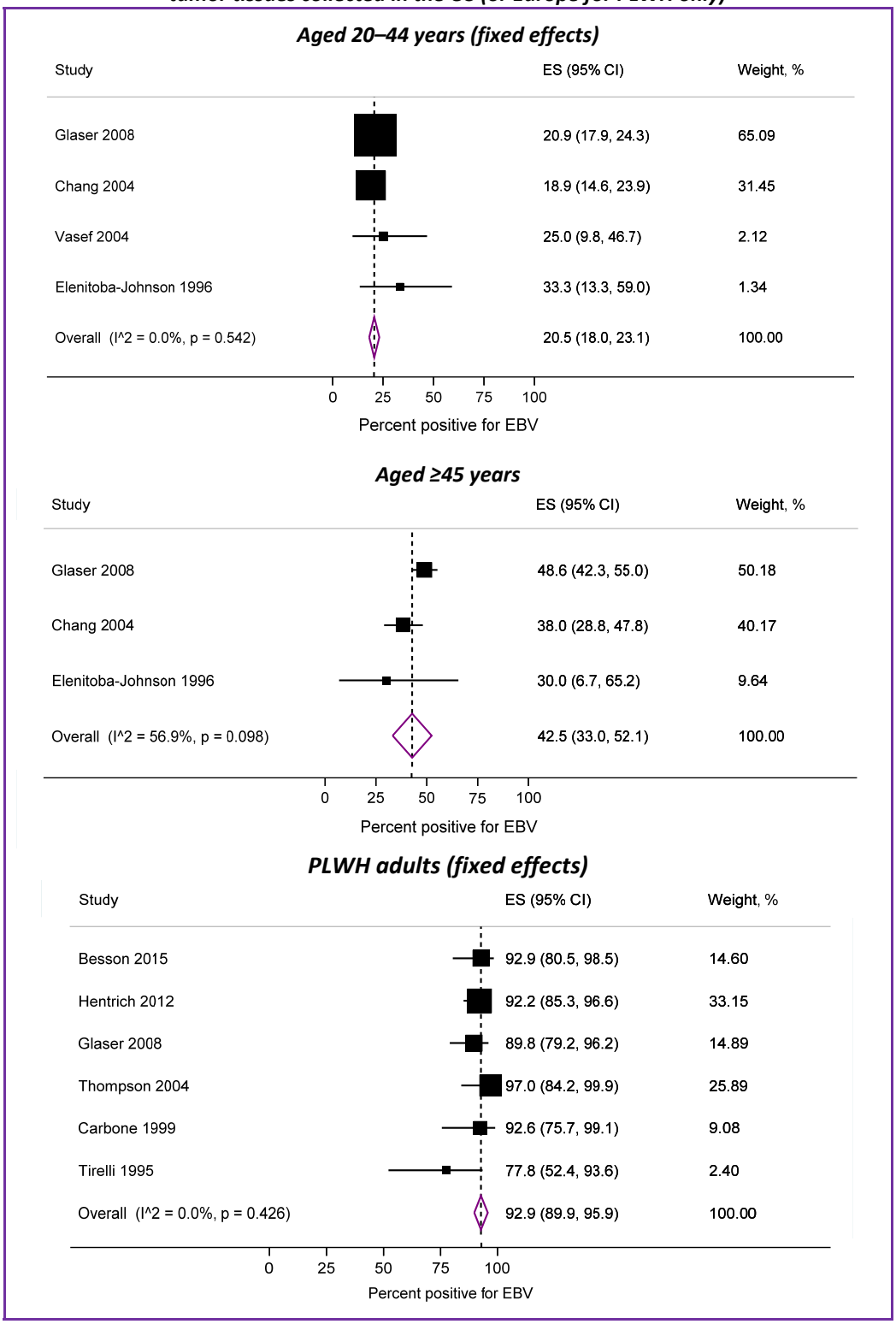
b. 26 cases from civilian sources, 15 cases from Veterans Administration medical centers, four cases from military hospitals.

Fig. S7. Forest plot of EBV prevalence (%) in Hodgkin lymphoma tumor tissues collected from individuals aged 0–19 in the US



CI = confidence interval, EBV = Epstein-Barr virus, ES = effect size, I^2 = index of consistency, US = United States

Fig. S8. Forest plot of EBV prevalence (%) in Hodgkin lymphoma tumor tissues collected in the US (or Europe for PLWH only)



CI = confidence interval, EBV = Epstein-Barr virus, ES = effect size, HIV = human immunodeficiency virus, I^2 = index of consistency, PLWH = people living with HIV, US = United States

Nasopharyngeal carcinoma (NPC)

A tumor of the epithelial tissues, NPC, is classified into three main types, keratinizing squamous cell accounting for 20% of all NPCs and non-keratinizing type accounting for the remaining 80% (further divided into differentiated and undifferentiated).[150] The pooled prevalence of the seven included studies reporting on EBV in NPC was 61.2% (CI: 45.1–77.2%) (**Table S14**, **Fig. S9**). For individuals aged 0–19 years old, we identified two eligible studies (Table S14)[319,320] each with eight NPC cases – all EBV positive. Since this was insufficient to perform a meta-analysis, we calculated exact CIs in OpenEpi[234] using a numerator and denominator of eight patients for a prevalence of 100.0% (CI: 63.1–100.0%).

Table S14. Characteristics of studies reporting on EBV prevalence in NPC cases

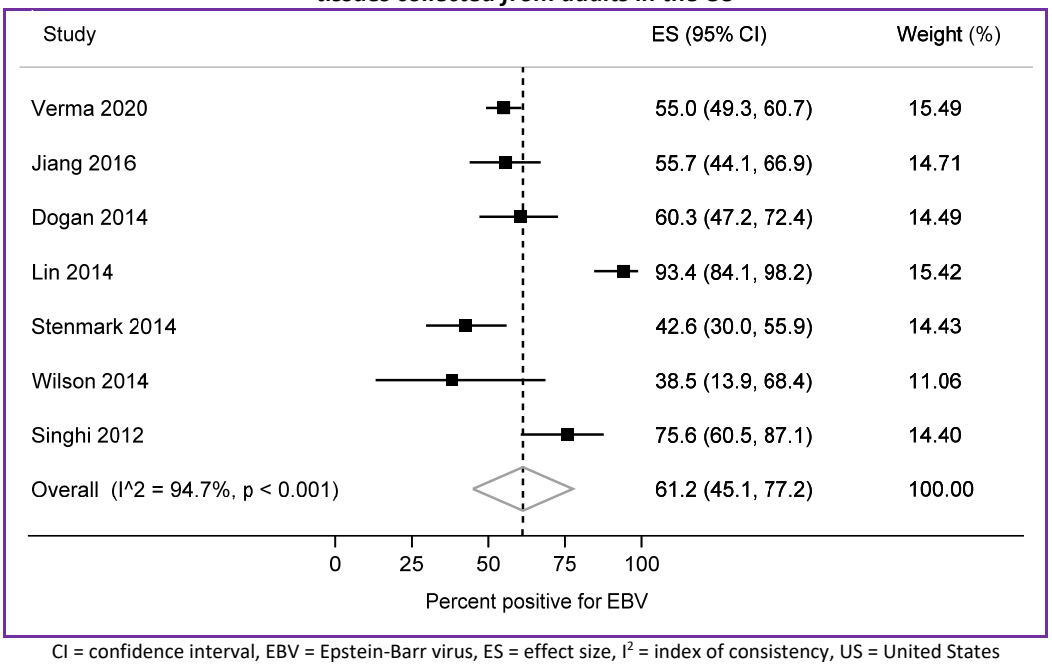
Study[a]	Region	Source of cases	Diagnosis dates	Male %	Mean/ median age	Non-keratinizing %	Tested n/N	Pos %
Studies conducted among adults (aged ≥20 years)								
Verma 2020[321]	New York	Memorial Sloan-Kettering Cancer Center	1998–2017	72.0	52.0	86.0	169/307	55.0
Jiang 2016[322]	Texas	M.D. Anderson Cancer Center	2000–2014	70.9	51.4	79.7	44/79	55.7
Dogan 2014[151]	Pennsylvania & Washington	University of Pittsburgh Medical Center, Virginia Mason Medical Center	1981–2012	69.8	53.0	85.7	38/63	60.3
Lin 2014[323]	California	Stanford University	1993–2010	75.4	45	NS	57/61	93.4
Stenmark 2014[324]	Michigan	University of Michigan	1985–2011	65.6	54.3	72.1	26/61	42.6
Wilson 2014[325]	Virginia	University of Virginia	2002–2013	NS	NS	76.9	5/13	38.5
Singhi 2012[326]	Maryland	Johns Hopkins Hospital	1985–2010	80.0	42.0	100.0	34/45	75.6
Studies conducted among children (aged 0–19 years)								
Polychronopoulou 2004[320]	Greece	Aghia Sophia Children's Hospital	1987–2001	NS	NS	100.0	8/8	100.0
Mertens 1997[319]	Germany	Institute of Pathology, University of Kiel	1992–NS	NS	NS	100.0	8/8	100.0

EBER ISH = Epstein-Barr encoding region *in situ* hybridization, EBV = Epstein-Barr virus, NPC = nasopharyngeal carcinoma, NS = not specified, US = United States

a. Inclusion criteria: tissue specimens from 10 of more cases tested for EBV, EBER ISH detection, conducted in the US, cases aged 15 and older.

b. EBV positivity was reported for two periods of diagnoses: 1956–1977 and 1981–2012, cases from the first diagnoses period were excluded because they occurred 61 to 40 years before year PAFs were applied to (2017).

Fig. S9. Forest plot of EBV prevalence (%) in nasopharyngeal carcinoma tumor tissues collected from adults in the US



CI = confidence interval, EBV = Epstein-Barr virus, ES = effect size, I^2 = index of consistency, US = United States

Extranodal natural killer T-cell lymphoma (ENKTL) – nasal type

EBV is detected in virtually all cases of ENKTL – nasal type and considered part of the diagnostic criteria for that cancer.[24,327-329] A study conducted at The University of Texas M.D. Anderson Cancer Center, reported that all 73 ENKTL – nasal type cases identified and tested, were EBER ISH positive.[328] All 186 ENKTL – nasal type cases were attributed to EBV.

Diffuse large B-cell lymphoma (DLBCL)

DLBCL, the most common subtype of NHL, has an average age of onset of mid-60s.[219] Studies meeting the inclusion criteria were published from 1996–2021, and all but one study reported on the HIV or the general immune status of cases thereby allowing us to calculate separate PAFs by HIV status (**Table S15**). Pooling 13 studies conducted in HIV negative populations and one study (Naeini 2016,[267] where HIV status was not reported) yielded EBV prevalence of 4.9% (**Fig. 10**). The pooled prevalence of EBV in DLBCLs diagnosed among PLWH was substantially higher at 45.7%. Utilizing estimated proportions of DLBCLs occurring in males with HIV (10.4% among those aged 0–29, 15.7% among those 30–59),[237] we partitioned the cancer incidence data and applied the pooled PAFs (4.9% and then 45.7% for PLWH).

Table S15. Characteristics of studies conducted in Canada, Europe or the US, reporting on EBV prevalence in DLBCL cases

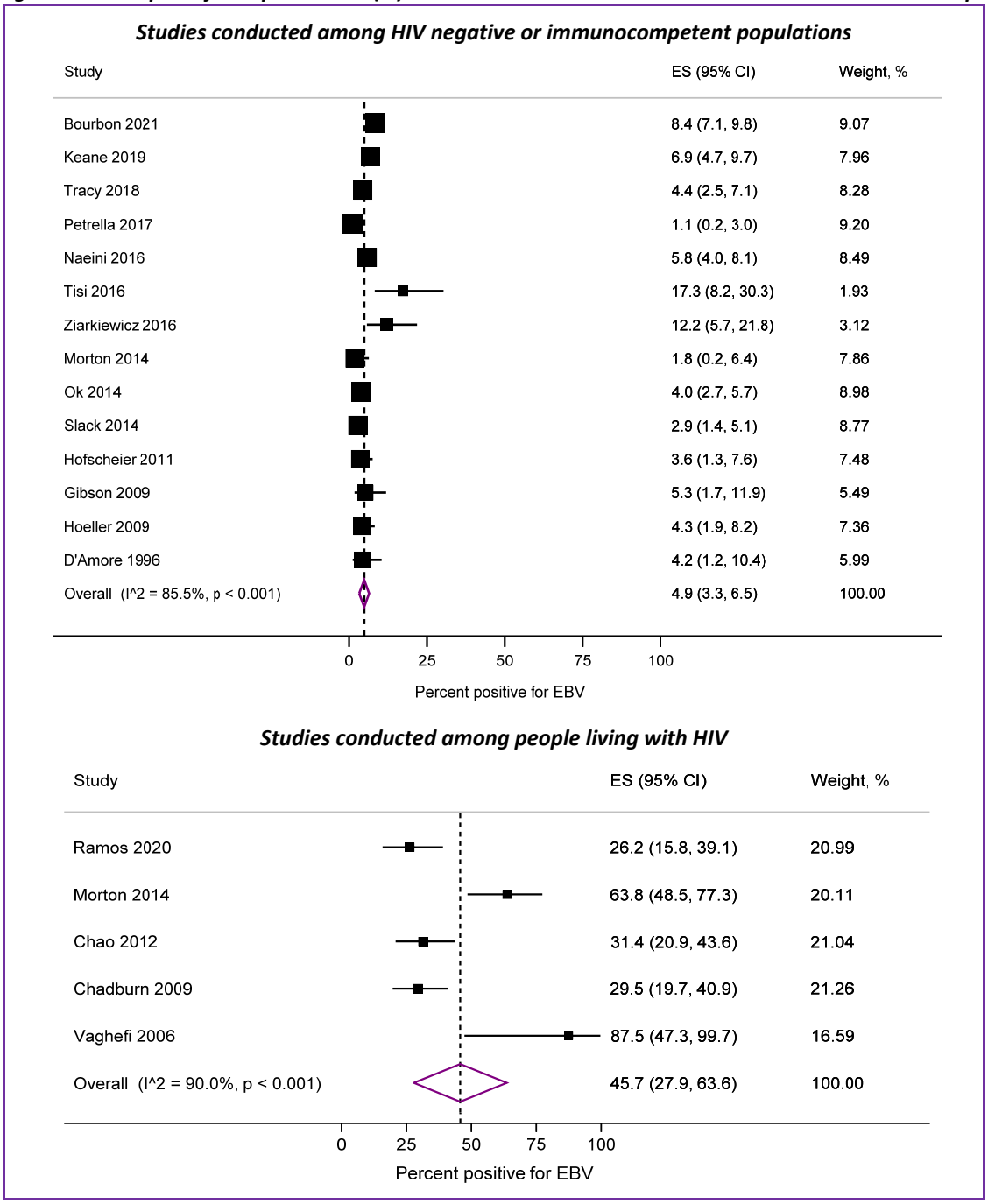
Study[a]	Region(s)	Source of cases	Diagnosis dates	Male %	Mean/ median age (range)	HIV status[b]	EBV positivity cut-off, %	Tested n/N	Pos %
Studies conducted among immunocompetent or HIV negative populations									
Bourbon 2021[330]	France	Hematopathology Department of Lyon-Sud University Hospital	2006–2019	NS	NS	IC	NS	138/1645	8.4
Keane 2019[331]	Australia & New Zealand	Princess Alexandra Hospital, Canberra Hospital, Royal North Shore Hospital, Australasian Leukaemia and Lymphoma Group Discovery Centre	2003–2014	40.0	NS (18–90)	IC	NS	30/433	6.9
Tracy 2018[332]	Iowa & Minnesota	University of Iowa or Mayo Clinic Rochester	2002–2012	59.4	63 (20–89)	HIV-	≥30	16/362	4.4
Petrella 2017[333]	Belgium, France, Switzerland	Lymphoma Study Association trial LNH03-6B	2003–2012	68.4	70 (60–80)	HIV-	NS	3/285	1.1
Naeini 2016[267]	California	Clarient Pathology Services/ Neogenomics	2008–2015	57.0	67 (11–96)	NS	≥10	33/567	5.8
Tisi 2016[334]	Italy	Catholic University of the Sacred Heart, Rome	2006–2013	NS	NS	HIV-	NS	9/52	17.3
Ziarkiewicz 2016[335]	Poland	Medical University of Warsaw	1994–2011	50.0	63.5 (23–86)	IC	>5	9/74	12.2
Morton 2014[336]	California	Los Angeles Residual Tissue Repository	1977–2003	48.3	NS	HIV-	"All or nearly all"	2/111	1.8
Ok 2014[337]	"Western countries"	International DLBCL Rituximab-CHOP Consortium Program Study	NS	57.5	63 (16–95)	HIV-	≥10	28/703	4.0
Slack 2014[338]	Canada	British Columbia Cancer Agency	1999–2006	63.3	64 (16–92)	IC	"Majority of tumor cells"	11/385	2.9
Hofscheier 2011[339]	Germany	Institute of Pathology, Tubingen	2000–2009	NS	72 (51–92)	IC	"Majority of tumor cells"	6/169	3.6
Gibson 2009[340]	Ohio	Department of Clinical Pathology, Cleveland Clinic	2002–2007	NS	NS (60–NS)	IC	NS	5/95	5.3
Hoeller 2009[341]	Austria, Italy & Switzerland	Pathology at the University Hospitals of Basel, Switzerland; Bologna, Italy; Innsbruck, Austria; and the Triemli Hospital, Zurich, Switzerland	NS	52.5	NS (50–93)	HIV-	≥10	8/188	4.3
D'Amore 1996[342]	Denmark	Danish Lymphoma Study Group (LYFO Registry)	1983–NS	NS	NS	IC	NS	4/95	4.2
Studies conducted among people living with HIV									
Ramos 2020[343]	US	Many study sites (Randomized controlled trial)	2012–2017	NS	NS	HIV+	NS	16/61	26.2
Morton 2014[212]	California	Los Angeles Residual Tissue Repository	1977–2003	100.0	NS	HIV+	"All or nearly all"	30/47	63.8
Chao 2012[344]	California	Kaiser Permanente Southern and Northern California Health Plans	1996–2007	91.4	NS	HIV+	≥75	22/70	31.4
Chadburn 2009[345]	California, Florida, Illinois, Massachusetts, New Jersey, New York, Ohio	Clinical trials AMC010 (45 pts) & AMC034 (36 pts)	NS	86.5	41	HIV+	"Majority of neoplastic cells"	23/78	29.5
Vaghefi 2006[346]	France	NS	1984–2002	NS	NS	HIV+	NS	7/8	87.5

DLBCL = diffuse large B-cell lymphoma, EBV = Epstein-Barr virus, HIV = human immunodeficiency virus, IC = immunocompetent, NS = not specified, Pos = positive, US = United States

a. Inclusion criteria: tissue specimens from 10 of more cases tested for EBV, EBER ISH detection, conducted in Canada, Europe or the US, cases aged 15 and older.

b. In addition to excluding HIV+ cases, some studies (reported as IC) made additional exclusions based on immune status (e.g., excluding organ transplant recipients).

Fig. S10. Forest plot of EBV prevalence (%) in DLBCL tumor tissues collected in North America or Europe



CI = confidence interval, DLBCL = diffuse large B-cell lymphoma, EBV = Epstein-Barr virus, ES = effect size, HIV = human immunodeficiency virus, I^2 = index of consistency

Gastric cancer

The association between EBV and gastric cancer was first reported in a case of lymphoepithelial-like gastric carcinoma,[347] and afterwards, the association was observed in gastric adenocarcinoma.[348] Since then, several meta-analyses have addressed the prevalence of EBV in gastric cancer prevalence.[349-353] The most recent systematic review by Tavakoli and colleagues including studies from 26 countries estimated a pooled prevalence of EBV infection (via EBER ISH detection) among gastric cancer patients of 8.77% (CI: 7.73–9.92).[354] Although the prevalence of EBV is higher in male than in female patients with gastric cancer, women are more likely than men to develop EBV-associated gastric cancer.[354]

We identified seven studies conducted in the US (**Table S16**), where the pooled prevalence of EBV was 1.9% for females and 13.6% for males (**Fig. S11**). We note that since we did not have data on EBV prevalence in DLBCL's occurring outside the stomach, we did not partition the estimates (EBV is associated with gastric cancer and DLBCL; rarely, DLBCL's can be diagnosed in the stomach [840 diagnoses in 2017]).

Table S16. Characteristics of studies reporting on EBV prevalence in gastric cancer cases in the US

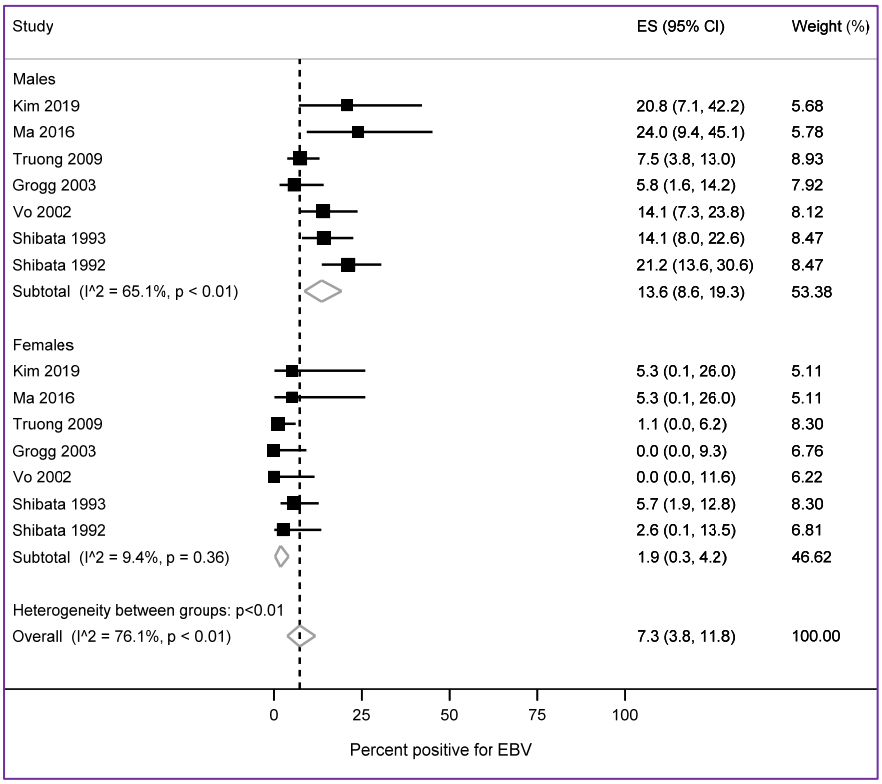
Study[a]	Region	Source of cases	Diagnosis dates	Mean/ median age in years	Males		Females	
					Tested n/N	Pos %	Tested n/N	Pos %
Kim 2019[355]	New York	Memorial Sloan Kettering Cancer Center	2006–2016	68.0 EBV+: 72.0	5/24	20.8	1/19	5.3
Ma 2016[356]	Pennsylvania	University of Pittsburgh Medical Center	2004–2015	73.0 EBV+: 68.0	6/25	24.0	1/19	5.3
Truong 2009[357]	Texas	University of Texas M. D. Anderson Cancer Center	1987–2006	EBV+: 60.0 EBV-: 67.0	11/147	7.5	1/88	1.1
Grogg 2003[358]	Minnesota	Mayo Clinic	1990–1998	68.4	4/69[b]	5.8	0/38	0.0
Vo 2002[359]	Texas, Louisiana, Minnesota	Touro Infirmary, St Luke's Baptist Hospital, Audie Murphy Memorial Veterans Administration Hospital	NS	EBV+: 66.5	11/78	14.1	0/30	0.0
Shibata 1993[360]	Hawaii	Japan-Hawaii Cancer Study	1965–NS	EBV+: 69.5 EBV-: 69.1	14/99	14.1	5/88	5.7
Shibata 1992[348]	Los Angeles	Hospital of the Good Samaritan, LAC+USC Medical Center	NS	NS	21/99	21.2	1/39	2.6

EBER ISH = Epstein-Barr encoding region *in situ* hybridization, EBV = Epstein-Barr virus, NS = not specified, Pos = positive, US = United States

a. Inclusion criteria: tissue specimens from 10 of more cases tested for EBV, EBER ISH detection, conducted in the US, cases aged 15 and older.

b. Removed three cases of known EBV-positive gastric carcinoma who were added to the series from the consultation files.

Fig. S11. Forest plot of EBV prevalence (%) in gastric cancer, by sex



CI = confidence interval, EBV = Epstein-Barr virus, ES = effect size

HUMAN PAPILLOMAVIRUS (HPV)

The most recent monograph (volume 100B) classified HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59 as Group 1 carcinogens.[1] HPV68 is considered 'probably' carcinogenic (Group 2A), and several HPV types (26, 30, 34, 53, 66, 67, 69, 70, 73, 82, 85, 97) as 'possibly' carcinogenic (Group 2B).[1] Persistent HPV infection is the strongest risk factor for anal, penile, vaginal and vulvar cancers, with virtually all cervical cancers being caused by HPV infection.[158]

Anal squamous cell carcinoma (SCC)

After combining six studies (**Table S17**) that met our inclusion criteria, the pooled prevalence of high-risk HPV (HR-HPV) in anal SCCs was 90.2% for males and 96.3% for females (**Fig. S12**). We found that 100% of anal SCCs among PLWH were attributable to HR-HPV. This finding is supported by studies conducted in Europe (not shown in Table 16); specifically, Kreuter and colleagues (2010) found HR-HPV in all nine HIV+ males diagnosed with anal SCCs from 2003 to 2009 in Germany;[361] Arana (2015) et al. reported that among 14 HIV+ males and five HIV+ females diagnosed with anal SCC in France from 2007 to 2009, all were HR-HPV+.[362] It has been estimated that 32.5% of anal SCCs in males and 3.0% in females were diagnosed in PLWH in the US from 2001 to 2015 (Shiels et al., 2022, *unpublished*

data). Only two of the five included studies reported HPV results by HIV status; among these two studies, 31.9% of cases were PLWH. We assumed that proportion of cases that are PLWH in studies where the HIV status of cases was not reported (Herfs 2017,[163] Alemany 2015,[160] Steinau 2013[164]) would be similar to that among the two studies (Zhu 2021[363] and Meyer 2013[162]) where HIV status was reported. For this reason, we combined all studies/cases to get PAFs for each males and females.

Table S17. Characteristics of studies reporting on HR-HPV prevalence in invasive anal SCCs in US populations, by sex and HIV status

Study[a]	Region(s)	Source of cases	Diagnosis dates	Histology	Detection methods HR-HPV types tested for[b]	Specimen	HIV status	Males		Females	
								Tested N	Pos %	Tested N	Pos %
Zhu 2021[363]	Massachusetts	Pathology archives	2000–2020	SCC	PCR, MGP, HPV GP5/GP6, L1 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68	FFPE	HIV-	34	64.7	70	88.6
							HIV+	12	100.0	0	NA
Herfs 2017[163]	Little Rock (Arkansas), Boston (Massachusetts)	Pathology archives	2001–2015	SCC	PCR-RT 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68	FFPE	Unknown (14/154 HIV+)	23	91.3	27	88.9
Alemany 2015[160]	Multiple	Pathology archives	1999–2009	SCC	SPF-10 PCR, DEIA, $LiPA_{25}$ 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73	FFPE	Unknown	35	88.6	57	100.0
Meyer 2013[162]	New York	Surgical pathology files	1997–2009	SCC	SPF-10 PCR, DEIA, $LiPA_{25}$ 16, 18, 31, 33, 35, 39, 45, 51, 56, 58, 59, 66, 68, 73	NS	HIV-	13	100.0	17	100.0
							HIV+	10	100.0	2	100.0
Steinau 2013[164]	Florida, Hawaii, Iowa, Kentucky, Louisiana, Michigan, California	Cancer registries, tissue repositories	1995–2005	133 SCC, 2 other[c]	PCR, LA, INNO-LiPA 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68	FFPE	Unknown	48	91.7	87	96.6

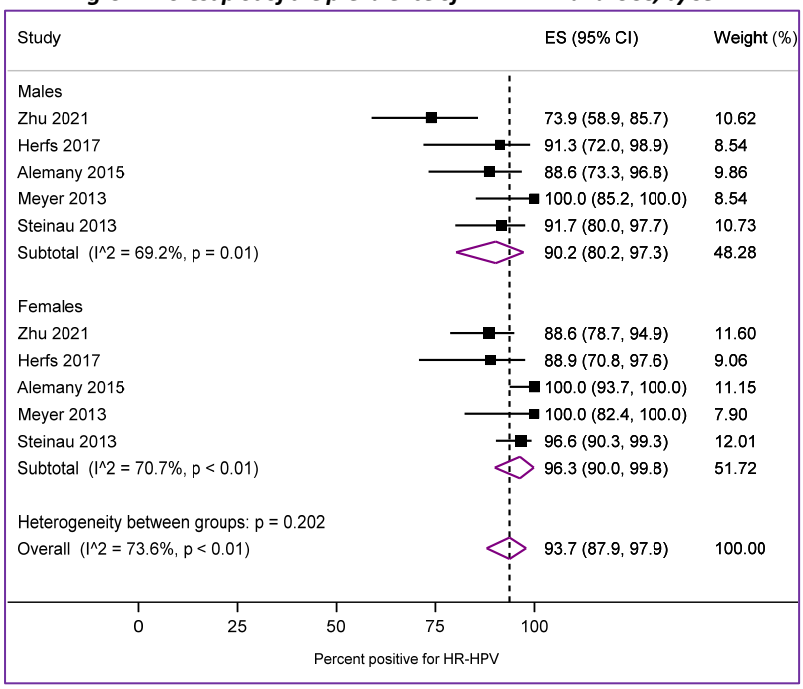
FFPE = formalin-fixed paraffin-embedded, HIV = human immunodeficiency virus, HPV = human papillomavirus, HR = high-risk, LA = Linear Array, MGP = modified general primer, NA = not applicable, NS = not specified, PCR = polymerase chain reaction, Pos = positive, RT = "RealTime", SCC = squamous cell carcinoma, US = United States

a. Inclusion criteria: invasive anal SCC tissue specimens, PCR detection, 10 or more cases, US-based study population, published after 1995, data stratified by sex or available upon request.

b. High-risk HPV types included: 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73 and 97.

c. We excluded 11 adenocarcinomas by removing five cases from males and six from females; 2/11 adenocarcinomas were HPV+ and one positive case was removed from each sex.

Fig. S12. Forest plot of the prevalence of HR-HPV in anal SCC, by sex[a]



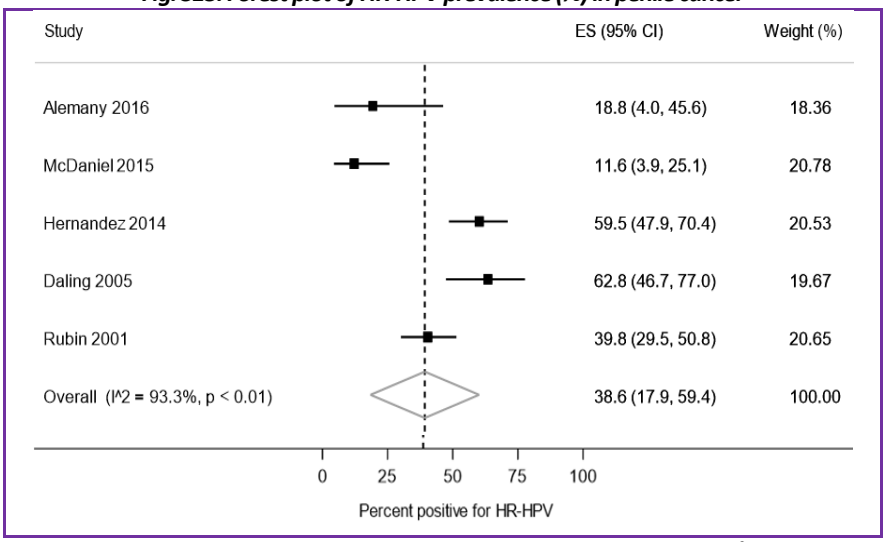
CI = confidence interval, HIV = human immunodeficiency virus, HPV = human papillomavirus, HR = high-risk, I^2 = index of consistency

[a.] The figures include those positive for HIV (e.g., Zhu 2021 included 34 HIV negative and 12 HIV positive males).

Penile cancer

Pooling the five studies (**Table S18** and **Fig. S13**) meeting the inclusion criteria provided a prevalence in cases of 38.6% (CI: 17.9–59.4%).

Fig. S13. Forest plot of HR-HPV prevalence (%) in penile cancer



CI = confidence interval, ES = effect size, HPV = human papillomavirus, HR = high-risk, I^2 = index of consistency

Table S18. Characteristics of studies reporting on HR-HPV prevalence in penile cancers diagnosed in the United States

Study	Region(s)	Source of cases	Diagnosis dates	Histology	Detection methods HR-HPV types genotyped[a]	Specimen	HIV status	Tested N	Pos %
Alemany 2016[165]	Hawaii[b], Iowa	Pathology archives	1994–2004	NS	SPF-10, DEIA, LiPA$_{25}$, 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59	FFPE	Unknown	16	18.8
McDaniel 2015[166]	Michigan	Pathology archives	2005–2013	SCC	GP5+/GP6+, MY09/MY11, CP, 16, 33	FFPE	Unknown	43	11.6
Hernandez 2014[167]	California, Florida, Hawaii[b], Iowa, Kentucky, Louisiana, Michigan	Population-based cancer registries, residual tissue repositories	1998–2005	NS (majority SCC)	PCR, LA, INNO-LiPA$_{25}$, 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68	FFPE	Unknown	79	59.5
Daling 2005[168]	Washington	Population-based cancer registry	1979–1998	NS	PCR-MY09/MY11, L1, 16, 18, 31, 33, 35, 45	PE	Unknown	43	62.8
Rubin 2001[169]	Connecticut, Michigan, New York, Texas	Pathology archives	NS	SCC	PCR SPF-10, LiPA$_{25}$, 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, 70	FFPE	Unknown	88	39.8

CP = consensus primers, FFPE = formalin-fixed paraffin-embedded, HIV = human immunodeficiency virus, HPV = human papillomavirus, HR = high-risk, LA = linear array, NS = not specified, PCR = polymerase chain reaction, PE = paraffin-embedded, Pos = positive, SCC = squamous cell carcinoma

Inclusion criteria: invasive penile cancers tissue specimens, PCR detection, 10 or more cases, North American study population, published after 1995

a. HR-HPV types included: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, 70.

b. Three cases overlapped.

Vaginal cancer

Two studies met our inclusion criteria (**Table S19**). The pooled prevalence of HR-HPV types in invasive vaginal cancers was 72.2% (CI: 62.8–81.7%).

Table S19. Characteristics of studies reporting on HR-HPV[a] prevalence in vaginal cancers in US populations and results of pooled analysis

Study[b]	Region(s)	Source of cases	Diagnosis dates	Histology	HIV status	Detection methods HR-HPV types tested[a]	Specimen	Tested N	Pos % 95% CI	Weight %
Sinno 2014[171]	California, Florida, Hawaii, Kentucky, Louisiana, Iowa, Michigan	Population-based cancer registries, residual tissue repositories	1994–2005	NS (86% SCC)	Unknown	LA, INNO-LiPA 16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 67, 68, 70, 73, 82	FFPE	60	75.0 (62.1–85.3)	74.7
Daling 2002[172]	Washington	Population-based cancer registry	1981–1998	SCC	Unknown	PCR-L1, MY09/MY11 16, 18/45, 31	PE	25	64.0 (42.5–82.0)	25.3

FFPE = formalin-fixed paraffin-embedded, HIV = human immunodeficiency virus, HPV = human papillomavirus, HR = high-risk, LA = linear array, NS = not specified, PCR = polymerase chain reaction, PE= paraffin-embedded, Pos = positive, SCC = squamous cell carcinoma, US = United States

a. High-risk HPV types included: 16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 67, 68, 70, 73, 82.

b. Inclusion criteria: invasive vaginal cancer tissue specimens, PCR detection, 10 or more cases, North American study population, published after 1995.

Vulvar cancer

Since HPV is more prevalent in vulvar cancers diagnosed among younger women, and vulvar cancer incidence is higher among older women,[364] HR-HPV prevalence was analyzed by age group (**Table S20**). The pooled prevalence of HR-HPV in cases was 74.4% for women aged <50 years and 45.7% for women aged ≥50 years old (**Fig. S14**).

Table S20. Characteristics of studies reporting on the prevalence of HR- HPV in vulvar cancer cancers in the US or Canada, by age-group

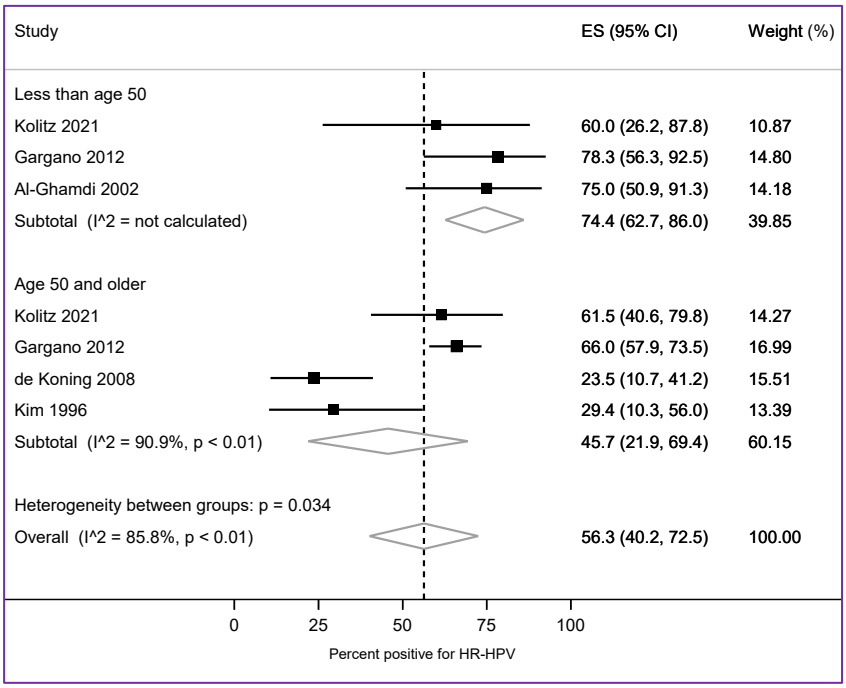
Study[a]	Region(s)	Source of cases	Diagnosis dates	Histology	HIV	Detection methods HR-HPV[b] types tested	Specimen	Age <50 years		Age ≥50 years	
								Tested N	Pos %	Tested N	Pos %
Kolitz 2021[365]	Texas	Pathology archives	2010–2020	SCC	None	Consensus PCR-L1, NS	FFPE	10	60.0	26	61.5
Gargano 2012[173]	California, Florida, Hawaii, Iowa Kentucky, Louisiana, Michigan	Population-based cancer registries, residual tissue repositories	1995–2005	NS	Unknown	LA, INNO-LiPA 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68	FFPE	23	78.3	153	66.0
de Koning 2008[174]	New York	Pathology department	1990–2005	SCC	Unknown	SPF-10, $LiPA_{25}$ 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, 70	PE	--		34	23.5
Al-Ghamdi 2002[175]	British Columbia, Yukon, Canada	Population-based cancer registry	1970–1998	SCC	One HIV+	PCR-MY09/MY11, PCR-GP5/GP6, TS 16, 18	FFPE	20	75.0	--	
Kim 1996[176]	Maryland, Florida	Pathology archives	1989–1994	SCC	Unknown	PCR-MY09/MY11, PCR-L1, TS, Sequencing 16, 18	Fresh	--		17	29.4

FFPE = formalin-fixed paraffin-embedded, HIV = human immunodeficiency virus, HPV = human papillomavirus, HR = high-risk, LA = linear array, NS = not specified, PE = paraffin-embedded, PCR = polymerase chain reaction, SCC = squamous cell carcinoma, US = United States

a. Inclusion criteria: invasive vulvar cancers tissue specimens, PCR detection, 10 or more cases, North American study population, published after 1995, data stratified by age or available upon request.

b. High-risk HPV types included: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, 70.

Fig. S14. Forest plot for HR- HPV prevalence in vulvar cancer, by age group



CI = confidence interval, ES = effect size, HR-HPV = high-risk human papillomaviruses, I^2 = index of consistency

Head and neck cancers (HNCs)

When attributing HNCs to HPV detection of the oncoproteins E6 and E7 is recognized as the gold standard,[177,178] because they are produced by HR-HPVs and must be present for viral replication to occur. We only considered the prevalence of HPV16 as the association between HNCs and HPV is most established for this type. Twenty-one studies met the inclusion criteria (**Table S21**). The PAFs were 60.3% for the oropharynx (**Fig. 15**), 7.9% for the oral cavity (**Fig. 16**) and 12.7% for the larynx (**Fig. 17**). Heterogeneity was high across the three sites.

Table S21. Characteristics of studies reporting on HPV16 prevalence detected via E6 and/or E7 in head and neck cancers in North American populations

Study[a]	Region[b]	Diagnosis dates	Detection method(s)	Speci-men	Anatomical site					
					Oropharynx		Oral cavity		Larynx	
					Tested N	Pos[c] %	Tested N	Pos[c] %	Tested N	Pos[c] %
Lewis 2021[366]	Tennessee	2000–2018	qRT-PCR E6/E7	FFPE	259	81.9	--		--	
Mazul 2016[367]	North Carolina	2002–2006	TS-PCR E7	FFPE	238	63.4	--		--	
Hooper 2015[180]	Oregon	--	PCR-E6, E7	FF	44	68.2	24	8.3	19	0.0
Zandberg 2015[368]	Maryland	1992–2007	PCR-E6	FFPE	194	34.5	--		--	
Isayeva 2014[181]	Alabama	2004–2012	RT-PCR E6/E7	PE	102	48.0	--		--	
Lingen 2013[183][d]	California, Illinois, Ohio, Ontario (CA)	2005–2011	qRT-PCR E6 or 7	FFPE	--		409	3.7	--	
Walline 2013[184]	Michigan	2001-2011	PCR-E6	FFPE	208	78.8	104	4.8	--	
Jordan 2012[185]	California, Illinois, Ohio, Ontario (CA)	2000–2009	qPCR E6	FFPE	235	62.1	--		--	
Stephen 2012[186]	Michigan	1999–2005	qRT-PCR E6	FFPE	--		--		77	27.3
Chaturvedi 2011[187]	Hawaii, Iowa, Los Angeles, California	1984–2004	qRT-PCR E6	FFPE	216	35.2	--		--	
Schlecht 2011[188]	New York	NS	TS-PCR E6/E7	FF, PE	23	52.2	29	27.6	27	18.5
Agoston 2010[189]	Massachu-setts	NS	PCR-E7	FFPE	126	58.7	--		--	
Kingma 2010[369]	Oklahoma & Montana	2005–2007	RT-PCR-E6	FFPE	61	49.2				
Jo 2009[190]	California	2000–2003	PCR-E7	FF, FFPE	14	92.9	--		--	
Settle 2009[191]	Maryland	1995–2006	PCR-E6	PE	--		28	10.7	55	7.3
Tezal 2009[192]	New York	1999–2005	TS-PCR E6	PE	30	70.0	--		--	
Cohen 2008[193]	Pennsylvania	1996–2001	TS-PCR E7	PE	35	68.6	--		--	
Liang 2008[194]	Minnesota	2004–2006	TS-PCR E6	FF	--		51	2.0	--	
Worden 2008[195]	Michigan	NS	RT-PCR E6	NS	42	64.3	--		--	
Zhao 2005[196]	Maryland	1984–2002	RT-PCR E6/E7	Frozen	26	57.7	38	15.8	16	18.8
Strome 2002[198]	Minnesota	1987–1995	TS-PCR E6	PE	52	40.4	--		--	

CA = Canada, FF = fresh-frozen, FFPE = formalin-fixed paraffin embedded, HPV = human papillomavirus, NS = not specified, PE = paraffin embedded, PCR = polymerase chain reaction, Pos = positive, qRT-PCR = real-time quantitative reverse transcription, RT = real-time, TS = type-specific, US = United states

-- Indicates the cancer was not included in the original study or that it overlapped with another included study.

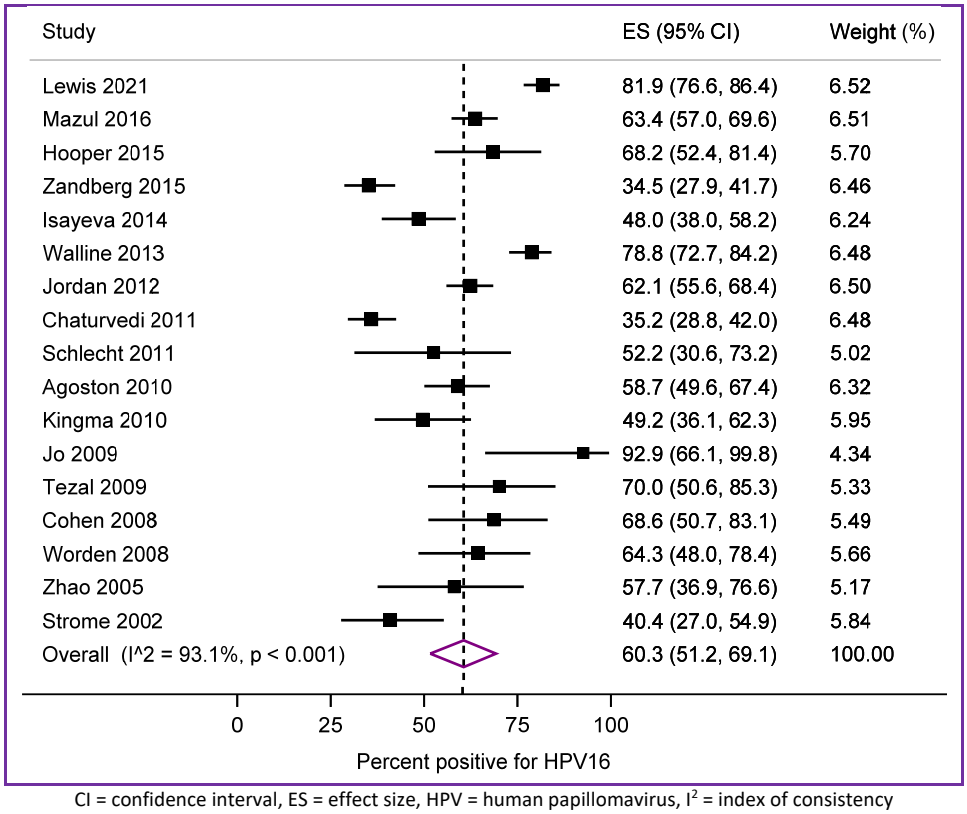
a. Inclusion criteria: site specific results (e.g., base versus oral tongue), detection in cancer tissue, invasive and untreated cancer, detection with E6 and/or E7 for HPV16, did not test specimens for E6/7 based on previous HPV results, North American study population, and published in 2000 or later.

b. Only cases from Chaturvedi et al.'s 2011 study originated from population-based cancer registries, the remaining studies cases came from clinics, hospitals, and pathology departments.

c. Tested positive for E6 and/or E7.

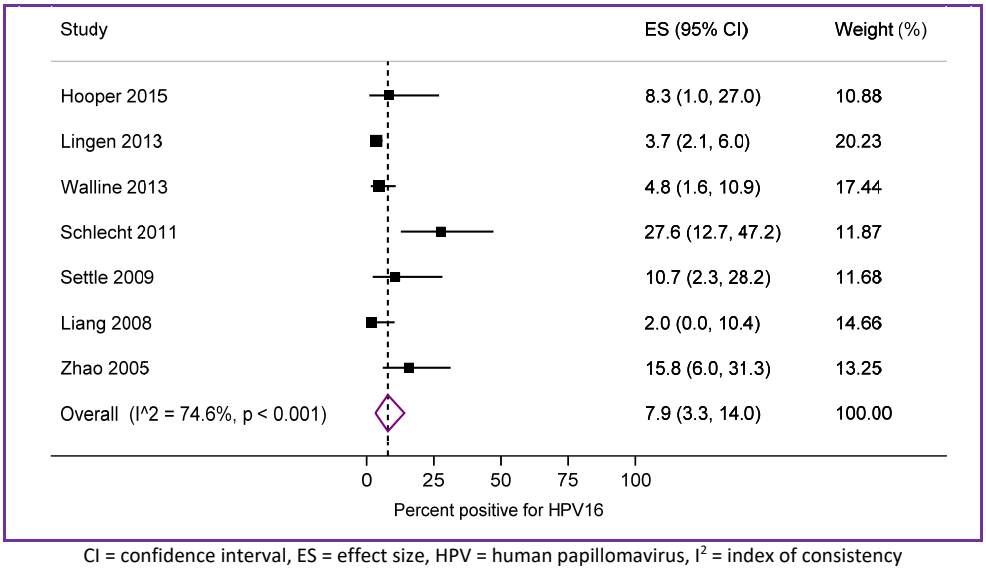
d. Lingen 2013 included four *in situ* cases.

Fig. S15. Forest plot of HPV16 E6/E7 prevalence in cancer of the oropharynx



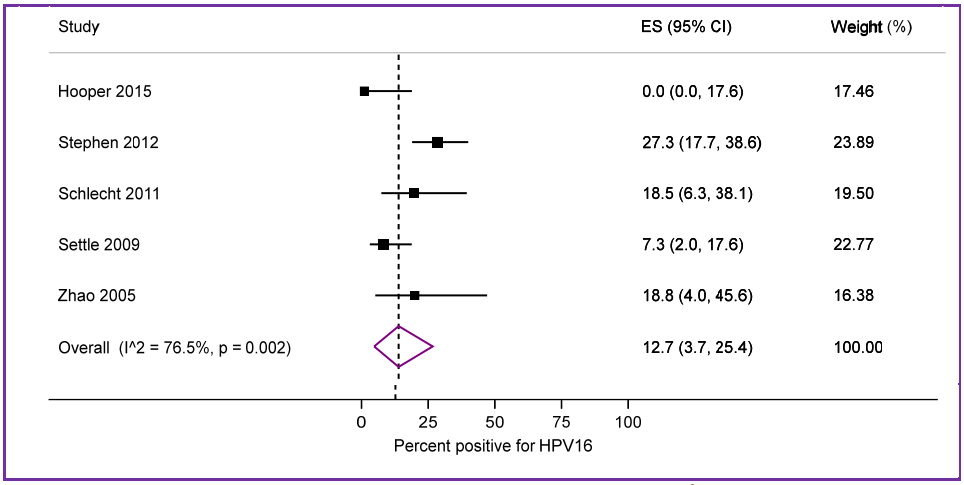
CI = confidence interval, ES = effect size, HPV = human papillomavirus, I^2 = index of consistency

Fig. S16. Forest plot of HPV16 E6/E7 prevalence in cancer of the oral cavity



CI = confidence interval, ES = effect size, HPV = human papillomavirus, I^2 = index of consistency

Fig. S17. Forest plot of HPV16 E6/E7 prevalence in cancer of the larynx



CI = confidence interval, ES = effect size, HPV = human papillomavirus, I^2 = index of consistency

Chapter 3: The Burden of Epstein-Barr Virus in Childhood and Early Adolescent Cancers

While numerous causes of cancers among adults have been identified, fewer risk factors for cancers occurring among children/adolescents have been discovered.[370,371] Although studies have quantified the role of EBV in associated cancers, this has not yet been accomplished for cancers occurring among children/adolescents.

Manuscript #3: Epstein-Barr virus and cancer among children and adolescents in Europe and North America: a systematic review, meta-analysis, and attributable burden estimation

This manuscript provides estimates of the impact that EBV has on Burkitt lymphoma, Hodgkin lymphoma and cancer of the nasopharynx among children/adolescents. While this thesis is focused on North America, for this study we also included studies conducted in Europe to enhance the overall value of the study and because we believe that EBV prevalence within cancers occurring in Europe would be comparable to that of North America.

This manuscript is not yet formatted for a specific journal.

Epstein-Barr virus and cancer among children and adolescents in Europe and North America: a systematic review, meta-analysis, and attributable burden estimation

Karena D. Volesky[1,2]
Olga Tsyruk[1]
Allan Hildesheim[3]
Stephen D. Walter[4]
Mariam El-Zein[1]
Christine M. Friedenreich[5,6]
Darren R. Brenner[6]
Eduardo L. Franco[1,2]

1. Division of Cancer Epidemiology, McGill University, Montreal, Canada
2. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada
3. Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, United States
4. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
5. Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada
6. Departments of Oncology and Community Health Sciences, University of Calgary, Calgary, Alberta, Canada

Financial support: This work is supported by a foundation grant (FDN-143347) to ELF and a doctoral award (support for KDV) jointly provided by Fonds de recherche du Québec – Santé (FRQS) and Fondation des étoiles.

Acknowledgements: We are grateful to Khandideh Williams for hand searching monographs and reviews for potential records, Callum JR Mullen for his assistance in managing the data, and Liaison Librarian at McGill University Genevieve Gore for assisting us in the preparation of the search strategy. We are also thankful that several researchers replied to our data requests – those who provided data are acknowledged in the first page of the supplement.

ABSTRACT

Background: Epstein-Barr virus (EBV) is a common infection in childhood and a cause of Burkitt lymphoma (BL), Hodgkin lymphoma (HL), and nasopharyngeal cancers (NPC). We estimated the percentage and number of incident BLs, HLs, and NPCs attributable to EBV among individuals aged 0 to 19 in Europe and North America in 2020.

Methods: A systematic review and meta-analysis was conducted by searching Embase and MEDLINE on July 27th, 2021, to identify studies where cancer tissues from at least eight individuals aged 0 to 19 years were tested for EBV by EBV-encoded RNA *in situ* hybridization or for HL EBV latent membrane protein 1 (LMP-1). Pooled prevalence of EBV in cancer tissues and its exact 95% confidence intervals (CIs) were calculated. To obtain the number of cases attributed to EBV, pooled proportions were multiplied by 2020 cancer incidence data derived from GLOBOCAN and the International Incidence of Childhood Cancer Volume III.

Results: The titles/abstracts of 13,818 records were screened and following full text reviews of 1,375 records, 49 studies met the inclusion criteria for BL (7 studies, 397 cases), HL (40 studies, 2,720 cases), and NPC (2 studies, 16 cases). The pooled prevalence of EBV was 15.5% (CI: 8.1–23.0%, I^2 = 70.4%) in BL, 37.9% (CI: 31.3–44.7%, I^2 = 90.0%) in HL, and 100% in NPC (16 cases). EBV prevalence in HL varied by several subgroup variables, including sex (males: 46.5% versus females: 23.7%, p = 0.007) and age (15–19 years: 15.3%, other age groups: 63.2–30.5%). Excluding studies published before the year 2000 and those not reporting EBV prevalence by age group, EBV prevalence in HL was re-pooled and applied to cancer incidence. An estimated 42,654 cancers were diagnosed in 2020, where 1,097 (2.6%, CI: 1.7–3.3%) were attributable to EBV. Of the 1,097 EBV attributable cases, 106 were BL, 837 HL, and 154 NPC. The burden of EBV attributable cancers was higher among males (3.0% of all cancers versus 2.1% for females) and among those aged 10–14 years (5.0% versus 0.8–2.7% for other age-groups).

Conclusion: A small proportion (2.6%) of cancers diagnosed among children/adolescents residing in Europe and North America is attributable to EBV.

1. INTRODUCTION

A cancer diagnosis in childhood or early adolescence is a life-changing, and sometimes life-taking, event. Currently, cancer is the second leading cause of mortality in children in Europe and the United States (US).[372,373] Among individuals aged 0 to 19 in Europe and North America, there were an estimated 42,654 cancer diagnoses and 6,104 cancer deaths in 2020.[374] While numerous causes of cancers among adults have been identified, fewer risk factors for cancers diagnosed in childhood have been discovered.[370,371] The Epstein-Barr virus (EBV) is among the few established causes of childhood cancers and is a common infection in children that can lead to certain lymphomas and epithelial tumours such as Burkitt lymphoma (BL), Hodgkin lymphoma (HL), and nasopharyngeal cancer (NPC).[1,371] Specifically, BL, a B-cell non-Hodgkin lymphoma (NHL), is an aggressive cancer that comprises a larger share of childhood cancers than adulthood cancers.[375] Both HL and NPC have bimodal age-specific incidence with peaks in adolescence and in the elderly.[376,377]

Globally, de Martel and colleagues estimated that EBV was responsible for 6,600 incident BLs, 40,000 HLs, and 110,000 NPCs in 2018 across all age groups;[57] specific estimates for children/adolescents were not provided. In fact, no published study has fully and specifically quantified the impact of EBV on childhood/adolescent cancer incidence. Estimates of the impact of EBV on childhood/early adolescent cancer incidence may be useful in the prioritization of vaccine development and identification of biological markers for early cancer detection.

The objective of this systematic review and meta-analysis is to provide estimates of the percentage and number of incident BLs, HLs, NPCs, and all cancers that are attributable to EBV among individuals aged 0 to 19 in Europe and North America in 2020. A secondary objective is to identify studies reporting on EBV prevalence in cancers other than BL, HL and NPC.

2. METHODS

Population attributable fractions (PAFs) quantify the percentage of cancer incidence that could have been avoided if EBV had been eliminated (e.g., EBV is prevented altogether or EBV infection is treated before cancer development) from the population. There is mechanistic evidence in support of the notion that when the EBV viral genome is detected within tumor cells (i.e., where its genome is not passively present in tumor cells but is transcribed and translated),

the cancer can be attributed to the EBV.[34,54] Therefore, consistent with previous work, the PAFs for BL, HL, and NPC for which the EBV is an established cause will be approximated by pooling studies that provided data on the prevalence of the EBV detected in cancer tissues.[57,67,139] To gather and select these studies in a reproducible manner, we performed a systematic review and meta-analysis and adhered to PRISMA guidelines in our reporting.[378] The protocol for this review was registered with PROSPERO (CRD42021269730).

EBV is also a cause of extranodal natural killer T-cell lymphoma (ENKTL) – nasal type, and detected in virtually all cases.[1] One study of six pediatric cases found EBV, as detected by EBV-encoded RNA *in situ* hybridization (EBER), in all cases.[379] This cancer is extremely rare in Europe and North America, representing less than 0.1% of NHLs.[143] Given its rarity and the lack of related cancer incidence data, we did not include ENKTL – nasal type in our review.

2.1. Search strategy

A systematic search was run in Embase (1947–) and MEDLINE (1946–) electronic databases on July 27th 2021, using a combination of Medical Subject Headings and keywords for the population of interest (i.e., infant, child, youth, adolescent, etc.), exposure (i.e., Epstein-Barr virus, herpesvirus 4, etc.) and outcome (i.e., neoplasms, cancer, lymphoma, etc.), without date restrictions but restricted to records published in English. The search was reviewed by a health sciences librarian with expertise in systematic search methods. The International Agency for Research on Cancer (IARC) monographs evaluating EBV (1997, 2012),[1,85] abstracts from relevant conferences occurring from 2010 to July 2021 (shown in **Appendix B**), and the reference lists of included records, were also searched for additional records.

2.2. Eligibility criteria

Any study, with any design that had at least eight cancer cases aged 0–19 with valid EBV testing was considered eligible. To be included, the cancer had to be invasive, primary, and not yet treated. When studies did not explicitly report whether cancer treatment had commenced at the time the cancer tissue specimen was taken, it was assumed that the cancer was not yet treated. Recurrent cases were excluded. Studies had to enroll patients whose specimens were collected in Canada, the US (including Puerto Rico), or Europe (including Cyprus, Greenland, Iceland, and Russia). Eligible studies had to test cancer tissues (i.e., biopsy or surgical specimens)

for EBV; studies only testing for EBV in sera or saliva were excluded. Additionally, the EBV must have been tested for with EBER1 or EBER2 ISH detection because it is considered the gold standard and the most reliable assay to detect EBV in cancer tissues;[120,380] studies utilizing LMP-1 immunohistochemistry (IHC) were also eligible for HL because LMP-1 is comparable to EBER ISH for detecting EBV in HL.[120] For studies that presented results using both detection methods, specimens testing positive by either method were considered EBV-positive.

For the HL/BL-PAF calculations, we included studies published in the year 2000 or later. This criterion was applied because the PAFs will be applied to cancers diagnosed in 2020 or later and restricting to studies published in the last ~20 years can help minimize temporal trends that could affect the fraction of BLs/HLs attributed to EBV. However, this criterion was not applied to studies on NPC because the number of eligible studies and cancer cases was so few that this criterion was unnecessary. Additionally, for HL, EBV prevalence had to be reported (or available upon request) for at least one of the following age-groups: 0–4, 5–9, 0–9, 10–14, 15–19 years to be included because both EBV prevalence and HL incidence vary by age.[145,374,377]

2.3. Data selection and extraction

One reviewer (KDV) performed the initial review (i.e., title/abstract screening) in Rayyan and the full-text review in EndNote.[381,382] Data were extracted by OT and verified by KDV. Data extracted included information on: study design; where patients were recruited; patient characteristics (major inclusion/exclusion criteria, how they were enrolled, immune status, cancer types and subtypes, ethnicity, sex, and age including mean, median, and range); specimens (collection date, type and preservation method); cancer assessment (how cancers were diagnosed); EBV detection method(s); the number of individuals tested; and number testing EBV positive overall and by subgroups (sex, age, HL subtypes, and when available, the joint distributions of these subgroup variables). When study populations overlapped, the record with the highest number of cases overall was retained, except if one study reported EBV prevalence with more granularity (e.g., by sex and/or age group). When the information required was not directly available in the publication, authors were contacted up to three times. Authors of records that were potentially eligible (i.e., the reported age range included several years in the eligible age range) were also contacted to clarify eligibility.

2.4. Quality assessment

Records were appraised with a modified version of the nine question/item Joanna Briggs Institute (JBI) Prevalence Critical Appraisal Tool.[383] Of the nine items, one (valid methods used for the identification of condition – EBV) was applied as an inclusion criterion and two others (adequacy of the response rate and appropriateness of statistical analysis) were not relevant. The remaining seven items were modified and applied as described in **Appendix C.** For example, for the tool's question: "Were study participants sampled in an appropriate way?", data on how cases were enrolled (sampled via consecutive, random, convenience, not reported) were captured. Additionally, for the question "Were the study subjects and the setting described in detail?" we captured two items – reporting the sex distribution of the study population and clearly reporting that specimens were collected pre-treatment. While the appraisal tool was not originally designed to provide a score *per se*, we assigned a point for the each of the seven items to obtain a summary quality score.

2.5. Statistical analysis

EBV prevalence was estimated by dividing the number testing EBV-positive by the number with valid testing results and indeterminate results were excluded from the denominator. All meta-analyses were performed in Stata/SE 16 (StataCorp, College Station, TX, USA). The *metaprop* command was used to estimate pooled proportions and their exact (Clopper-Pearson) 95% CIs for each of BL and HL.[110,233] Individual studies were combined using a random effects model, where the pooled estimate was estimated with the DerSimonian and Laird method, and the Freeman-Tukey double arcsine transformation stabilized variance. Heterogeneity was examined via sub-group analyses, the index of consistency (I^2), and the Cochran's Q p-value. A fixed effect model was adopted if the index of consistency (I^2) was <25%, and the test for heterogeneity, Cochran Q test, was not statistically significant ($p>0.10$). For the PAF estimates, data for HL were re-analyzed using the same meta-analytic techniques just described.

Sensitivity analyses were conducted to assess how the pooled estimate was influenced by the removal of studies that were: conducted solely among nodular lymphocyte-predominant HL subtype (EBV is generally absent in nodular lymphocyte-predominant Hodgkin lymphoma [NLPHL] cases)[384,385], reported as abstracts or short communications, assessed to have quality

scores of less than four, or published before 2000. Subgroup analyses were conducted according to sex, age group (several age groups are presented because of the variability in how EBV prevalence was reported by age across studies), year of publication, region, enrollment method, number of cases, HL subtype, and EBV detection method. For studies to be meta-analyzed, there had to be at least eight participants in any subgroup; for example, a study reporting on EBV prevalence among 15 cases of HLs (11 males and four females) will be included, but the EBV results for its four females would not be included in the subgroup meta-analysis among females.

2.6. Cancer incidence

To estimate the number of cancers attributable to EBV, the PAF was multiplied by the number of incident cases. Estimates of HL (C81), NHL (C82–86, C96), cancer of the nasopharynx (C11), and all cancers (C00–97), were acquired from GLOBOCAN for the most recent year available, 2020.[374] BL cases were estimated by partitioning the GLOBOCAN NHL estimates by the proportion expected to be BL. To do so, the International Incidence of Childhood Cancer (IICC) Volume III data were used to calculate these proportions by sex and five-year age groups for each country.[386] For countries where GLOBOCAN data were available but IICC data were not (Albania, Bosnia and Herzegovina, Denmark, Finland, Hungary, Latvia, Moldova, Montenegro, North Macedonia, Romania, and Serbia), the proportion of NHL estimated to be BL was approximated by extrapolating the proportions from two or more nearby countries. **Supplementary Table 1** summarizes the specific methods used to estimate cancer incidence for each country. A sensitivity analysis was conducted to assess how excluding these countries influenced the total percentage of cancers attributable to EBV. Due to a lack of GLOBOCAN data, we did not calculate any estimates for Andorra, Greenland, Liechtenstein, Monaco, Kosovo, or San Marino.

3. RESULTS

The titles/abstracts of 13,818 records were screened, of which 1,375 records underwent full-text review (**Fig. 1**). Fifty-two studies met the inclusion criteria, three[298,299,387] of which reported on more than one cancer type. Seven studies (397 cases) reported on BL,[296-298,300,301,388,389] 40 studies (2,720 cases) on HL,[146,147,312-318,387,390-419] two studies (16 cases) on NPC,[319,320] and five on other cancers.[299,387,409,420,421] Of the 49 studies reporting on BL, HL, and NPC, 25 were conducted exclusively in pediatric/adolescent populations. A summary of the characteristics of the individual studies included for BL, HL, and NPC can be found in **Supplementary Tables 2**, **3**, and **4**, respectively.

3.1. Burkitt lymphoma

The seven studies reporting on BL were published from 2004 to 2021 and enrolled between 13 and 222 cases and reported EBV prevalence ranging from 5.6% to 57.1% (Supplementary Table 2).[296,301] The pooled prevalence of EBV in BL was 15.5% (CI: 8.1–23.0%) and heterogeneity was high (I^2 = 70.4%) (**Fig. 2**).

3.2. Hodgkin lymphoma

The 40 included studies were published from 1992 to 2021, 70% were conducted in Europe, and enrolled a median of 28 cases (range: 8–842) (Supplementary Table 3). Among 2,720 HL cases with valid testing, 853 (31.4%) tested positive for EBV – pooling these data with random effects produced a pooled prevalence of 37.9% (CI: 31.3–44.7%) (**Fig. 3**). A summary of the sensitivity and subgroup analyses are presented in Table 1 (accompanying subgroup forest plots are in **Supplementary Figs. 1**–**8**). While the subgroup analyses revealed several variables influencing EBV prevalence, the more notable differences were by age group (adolescents aged 15 to 19 had much lower EBV prevalence compared to other age-groups), and HL subtype (79.5% of mixed cellularity HL cases were EBV positive versus 7.6% of NLPHL cases). There was considerable heterogeneity within and between subgroups. Most I^2 values within subgroups exceeded 75%. While higher prevalence of EBV was found in studies conducted in Southern Europe (80.7%) compared to all other regions, we note that the five[409,414-416,419] studies enrolled few cases (13–24 cases), and three of these were conducted from 1994 to 1996.[409,414,419]

3.3. Nasopharyngeal cancer

Two studies[319,320] with eight cases each, all EBV positive, were insufficient to perform a meta-analysis of NPC (Table 4). Instead, exact CIs were calculated using an open source tool for one study[234] with a numerator and denominator of eight patients for a prevalence of 100.0% (CI: 63.1–100.0%).

3.4. Other cancers

Other cancers and their associated EBV prevalence (number EBV positive/number tested) were: anaplastic large-cell lymphoma (0/44[299] and 0/11),[387] diffuse large B-cell lymphoma (0/64),[299] lymphoblastic lymphoma – type B (5/10),[298] precursor B-lymphoblastic lymphoma (0/15),[299] peripheral T-cell lymphoma (0/11),[299] T-cell lymphoblastic lymphoma (0/19),[409] inflammatory myofibroblastic tumor (2/15),[421] and salivary gland (0/10).[420] While EBV is implicated in a subset of gastric carcinomas (~10%), we did not identify any eligible studies in this age range.

3.5. PAF analyses

Given the limited data on BL and NPC, PAFs could not be estimated by sex and/or age groups, thus a single PAF was applied to each BL and NPC cancer incidence. After excluding 18 studies published before the year 2000 and nine studies that did not report EBV prevalence for at least one age group (possible age-groups were: 0–4, 5–9, 0–9, 10–14, 15–19), there remained 13 studies for the HL PAF calculations.[147,312,314,315,392,393,395,398,399,401,404,407,415] As a sensitivity analysis, we estimated PAFs for HL by sex; there were 10 studies meeting the inclusion criteria reporting EBV prevalence by sex.[146,315,393,398,399,401,404,415,417] The data available by both sex and age group were too sparse to permit simultaneous stratification by sex and age.

Among the 1,097 cancers attributed to EBV (2.6% of all cancers), 9.7% were BL, 76.3% were HL, and 14.0% were NPC (Table 2). After applying the age-group PAFs to HL incidence, 25.0% (837/3,353) of HLs were attributed to EBV. Using sex-based PAFs for HL, instead of age-based ones, led to more cancers being attributed to EBV – 32.2% versus 25.0%. The percent of all cancers attributable to EBV was higher for males compared to females (3.0% versus 2.1%) and highest for those aged 10 to 14 (5.0% versus 0.8–2.7% for other age-groups). The percent of

cancers attributable to EBV was estimated for each country and varied from less than 1% to 4.6% (**Supplementary Table 1**).

4. DISCUSSION

By pooling EBV prevalence from 22 studies, we estimated that 2.6% of the projected 42,654 cancers diagnosed among individuals aged 0 to 19 in 2020 in Europe and North America were attributable to EBV. Among the three EBV-attributable cancers we considered, the majority (76.3%) were HLs. While EBV occupies a small role in total childhood/adolescent cancer incidence, considering how few *potentially* modifiable causes of childhood cancers have been identified (many identified causes are not modifiable because they are hereditary; hereditary factors are estimated to account for about 6–8% of all cancers in childhood),[422,423] this association presents an opportunity to develop prevention/early treatment of cancer in a group where few cancers have a known cause and for which cancer has lifelong consequences.[370,371]

The prevailing type of BL in Europe and North America is the sporadic type, for which fewer cases are linked to EBV, unlike the endemic type where almost all cases are linked to EBV.[385] Previously, 10–40% of sporadic BLs have been attributed to EBV.[1,54,424,425] Our estimate of 15.5% is consistent (albeit on the lower end) with previous work despite previous estimates including all ages.[1,54,424,425] The proportion of HL that is EBV-positive depends on the region, age, sex, disease subtype, and immune status.[24] In Europe and North America, EBV prevalence in HLs varies from 20–50% across all age groups.[425] We estimated that 25.0% of HLs among those aged 0 to 19 could be attributed to EBV. However, when sex-based PAFs were applied to cancer incidence, 32.2% of HLs were attributable to EBV. This difference in PAF values may be explained by HL cancer incidence varying by both sex and age. Specifically, using GLOBOCAN data for Europe and North America combined, the age standardized incidence rate per 100,000 for HL increased with age for both males (0–4: 0.24, 5–9: 0.64, 10–14: 1.8, 15–19: 2.7) and females (0–4: 0.05, 5–9: 0.25, 10–14: 1.7, 15–19: 2.7) while we found that EBV prevalence steadily declined by age group thereby resulting in fewer cases being attributed to EBV than would be with sex-based PAFs.

A previous systematic review and meta-analysis found that globally EBV positivity in HL was higher in children (69.7%) than adults (41.1%), and that EBV prevalence was higher amongst

males, for mixed cellularity subtype, and in low income regions.[145] The prevalence reported for children (69.7%) across all regions was notably higher than what was found here, but EBV prevalence tends to be lower in more developed countries. The virus is most often associated with classical HL,[1] in particular nodular sclerosis and mixed cellularity subtypes, the remaining, non-classical HL, is NLPHL which is rarely EBV-positive.[426] Although less prevalent in NLPHL, EBV was still present in a small proportion (7.6%) among children. In contrast, most mixed cellularity (79.5%) subtypes were EBV positive.

Globally, in adults, 80% of NPCs in low NPC incidence regions and 100% in high NPC incidence regions were attributed to EBV.[54] However, the 80% estimate for low incidence areas was based on very sparse data.[54] We found that based on merely two studies of eight cases each, EBV prevalence was 100%.[319,320] While not meeting the inclusion criteria of a minimum of eight cases, another study with seven NPC cases aged 9 to 23 reported that all seven cases were positive for EBV via EBER-ISH.[324] NPC is a rare cancer in Europe and North America (the highest incidence is seen in Southern China, Singapore and Malaysia)[54] and especially rare among those aged 0 to 19.

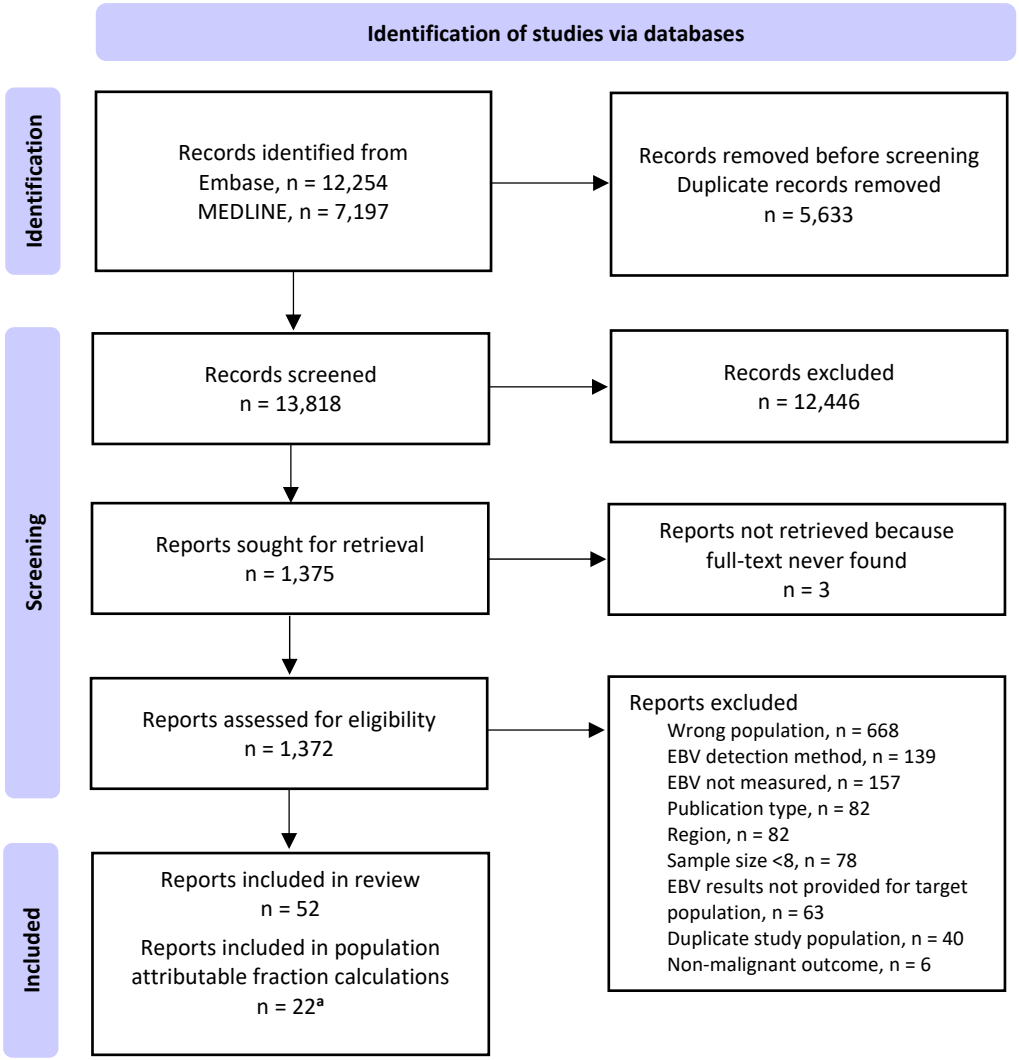
The main study limitations of our review are the lack of available evidence, and the quality of that evidence. We applied rigorous inclusion criteria to ensure that only studies meeting certain criteria, such as utilizing gold standard EBV detection methods, were included. While we restricted attention to studies conducted in Europe and North America, it does not preclude the possibility that patients could be referred from outside these regions. Of the 49 studies reporting on BL, HL, and NPC, only half (51.0%) were conducted exclusively in the target population (age 0–19). For this reason, the demographic characteristics presented for the larger study population including adults could not be ascertained for the pediatric/adolescent population. Because of limited data, we were not able to stratify by both sex and age simultaneously in the HL-PAF estimates – yet both factors influence EBV prevalence. We selected the factor that affected EBV prevalence the most (i.e., age) and utilized the other factor (sex) to assess the robustness of the resulting number of attributable cases. While the HL PAF analysis was based on studies published in 2000 or later (which included HL cases diagnosed before 2000) we cannot exclude the possibility of EBV prevalence in HL cases diagnosed in 2020 could differ from the prevalence

among cases diagnosed in earlier years. Uncertainty in the cancer incidence estimates was not accounted for (GLOBOCAN incidence estimates did not have CIs available). Furthermore, BL incidence had to be estimated from other data (IICC). When we excluded data from 11 countries where the proportion of NHL expected to be BL had to be extrapolated from other countries, we did not observe a meaningful change in the total percentage of cancers attributable to the EBV (2.571% versus 2.575% after removal of the 11 countries). Finally, the two studies used to estimate the PAF for NPC enrolled nasopharyngeal carcinoma cases. While nasopharyngeal carcinoma is the dominant histology in cancer of the nasopharynx, our application of PAF based on nasopharyngeal carcinoma to cancer of nasopharynx may slightly overestimate the number of cases attributable.

These findings cannot be extrapolated to other geographic areas. Despite how widespread and constant EBV infection is across the world's regions, EBV prevalence in BL, HL and NPC is known to vary by region. In addition, cofactors in the carcinogenesis of EBV related cancers, such as malaria and human immunodeficiency virus, are highly disparate across regions.[56,137,145,385,427]

In conclusion, our findings indicate that EBV was responsible for an estimated 15.5% of BLs, 25.0% of HLs, and 100% of NPCs in childhood/adolescent cancers diagnosed in Europe and North America in 2020.

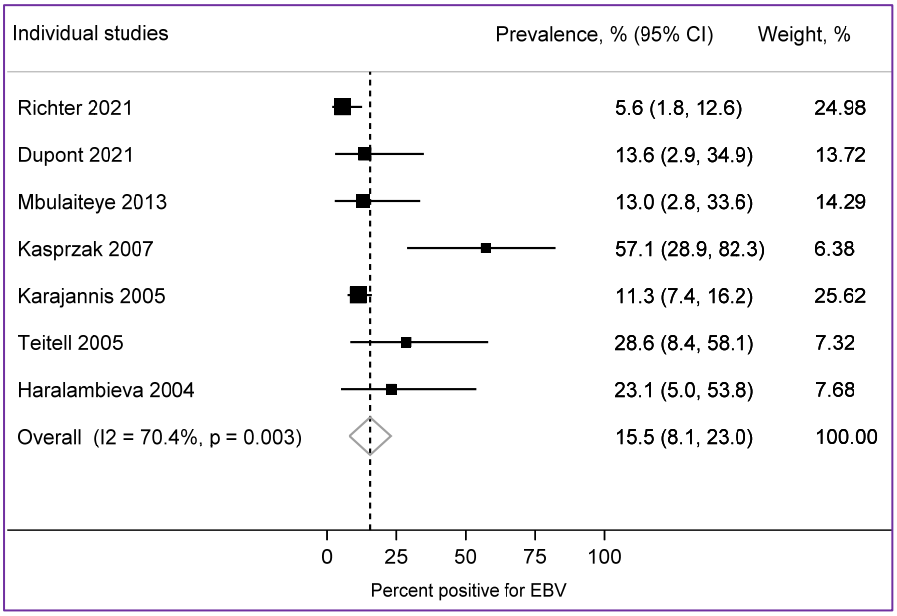
Fig. 1. Flow chart of search results and selection of studies examining EBV prevalence in cancer tissues



EBV = Epstein-Barr virus

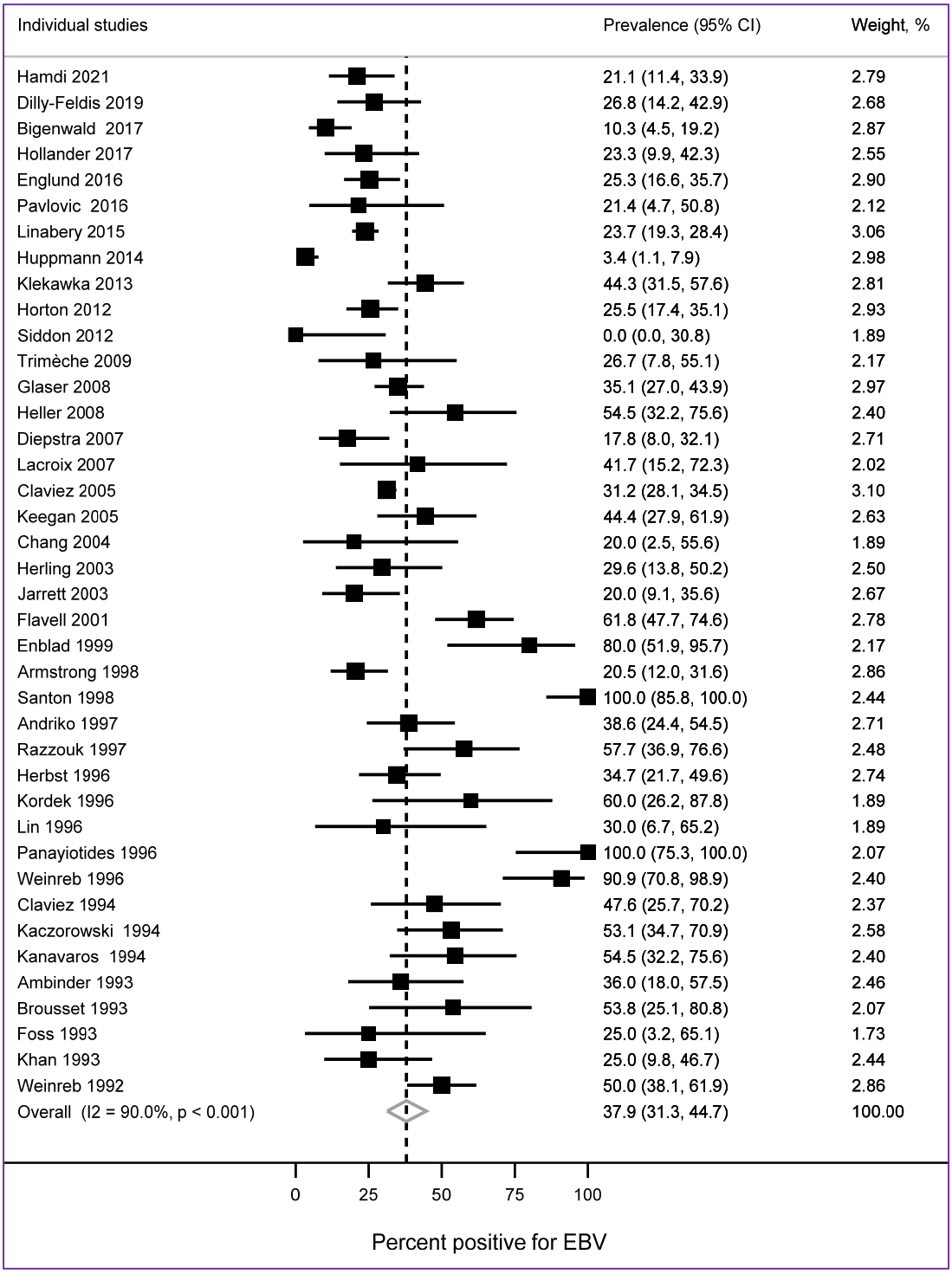
[a] Additional inclusion criteria were applied to select studies for the Burkitt lymphoma and Hodgkin lymphoma PAF analysis, these were: published in the year 2000 or later, and for HL only that EBV prevalence be reported for at least one of the following age-groups: 0–4, 5–9, 0–9, 10–14, 15–19.

Fig. 2. Forest plot of EBV prevalence (%) in Burkitt lymphoma tumor tissues collected from individuals aged 0–19 residing in Europe or North America



EBV = Epstein-Barr virus; CI = confidence interval

Fig. 3. Forest plot of EBV prevalence (%) in Hodgkin lymphoma tumor tissues collected from individuals aged 0–19 residing in Europe or North America



EBV = Epstein-Barr virus, CI = confidence interval

Table 1. Estimates of EBV prevalence in HL tumor tissues from sensitivity, subgroup, and PAF analyses

	Number of		Pooled prevalence, %[a]		Heterogeneity	
	Studies	Cases	Estimate	95% CI	I^2 (%)[a]	Cochran's Q p-value
Overall	40	2720	37.9	31.3–44.7	90.0	< 0.001
Sensitivity analyses (*by removing*)						
Exclusively NLPHL[b]	39	2575	39.2	33.1–45.5	87.2	<0.001
Abstracts/short communications	37	2572	37.0	30.1–44.2	90.3	<0.001
Quality score < 4[c]	25	2012	35.6	27.5–43.1	90.2	<0.001
Published before 2000	21	2215	25.3	19.3–31.8	86.8	<0.001
Sub-group analyses						
Sex						
Males	15	897	46.5	31.6–61.7	92.8	<0.001
Females	14	604	23.7	16.8–31.2	52.4	0.011
Age-group (years)[d]						
0 to 4	2	35	63.2	49.8–75.7	--	--
5 to 9	4	265	51.3	40.9–61.7	53.4	0.092
0 to 9	6	259	65.8	59.6–71.8	0.0	0.709
0 to 14	13	964	49.3	39.4–59.2	87.3	<0.001
10 to 14	9	221	30.5	20.8–41.0	83.5	<0.001
15 to 19	10	563	15.3	12.2–18.6	0.0	0.530
10 to 19	12	1268	30.8	21.4–41.0	77.6	<0.001
Year of publication						
2010 or later	11	980	19.4	12.1–27.8	87.0	<0.001
2000 to 2009	11	1235	34.6	27.1–42.2	72.3	<0.001
1990 to 1999	18	505	55.2	41.8–68.6	90.0	<0.001
Region[e]						
Western Europe	11	1181	27.3	20.3–35.0	69.1	<0.001
Eastern Europe	3	103	48.5	38.5–58.5	--	--
Northern Europe	8	398	36.6	23.4–50.8	86.9	<0.001
Southern Europe	5	95	80.7	46.1–100.0	91.7	<0.001
North America	12	916	28.7	18.4–40.2	89.8	<0.001
Method of enrolling cases						
Consecutive	20	1041	32.6	23.7–42.2	89.0	<0.001
Other/not specified	20	1679	43.9	33.9–54.1	90.6	<0.001
Number of cases						
8 to 19	11	130	42.1	21.4–64.2	83.0	<0.001
20 to 49	17	530	44.9	32.9–57.1	87.3	<0.001
50 to 99	7	485	32.0	19.0–46.6	90.9	<0.001
100 or more	5	1575	22.3	12.6–33.8	95.3	<0.001
HL subtype						
Nodular sclerosis	21	1152	32.0	22.9–41.8	85.8	<0.001
Mixed cellularity	13	388	79.5	71.2–86.8	51.8	0.015
NLP	7	328	7.6	2.0–15.5	68.3	0.004
Detection method						
EBER only	8	491	28.1	16.5–41.2	74.8	<0.001
LMP-1 only	15	1311	54.8	40.5–68.8	93.4	<0.001
EBER and/or LMP1	17	918	28.9	21.0–37.5	85.1	<0.001
Utilized in PAF calculations (years)[f]						
0 to 9	5	333	61.2	47.7–74.0	75.1	0.003
10 to 14	4	718	29.9	17.8–43.5	90.1	<0.001
15 to 19	9	506	15.0	11.8–18.5	0.0	0.437
Sensitivity analysis for PAF calculations						
Males	9	658	41.7	31.4–52.4	70.5	<0.001
Females	9	535	20.8	17.2–24.6	15.6	0.304

EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, NLPHL = nodular lymphocyte-predominant HL, CI = confidence interval, I^2 = index of consistency, EBER ISH = Epstein-Barr encoding region *in situ* hybridization; LMP-1 = latent membrane protein 1, PAF = population attributable fraction, -- not calculable

a. When the I^2 was <25%, a fixed effects model was used.

b. One study enrolled NLPHL cases only.[406]

c. Quality scored based on seven items (a study could be awarded a maximum of seven points); the quality assessment is described in appendix C.

d. Each group includes prevalence estimates for studies that presented 0 to 9 age group (e.g., age-groups 0–4 and 5–9 were not reclassified into the 0–9 group).

e. One study was conducted in both Europe and North America and was excluded from the region-based subgroup analysis.[403]

f. Excluded from the PAF analyses were: 18 studies published before 2000, five studies that did not report EBV positivity by either sex and/or age group, and one study that enrolled NLPHL cases only.

Table 2. Estimates of the percentage and number of cancers attributable to EBV among individuals aged 0 to 19 residing in Europe and North America

	All cancers					Burkitt lymphoma			Hodgkin lymphoma			Nasopharyngeal cancer		
	N	PAF %	95% CI	AC	AC 95% CI	N	AC	95% CI	N	AC	95% CI	N	AC	95% CI
Overall	42,654	2.6	1.7–3.3	1097	742–1419	680	106	56–157	3353	837	589–1108	154	154	97–154
Sensitivity analyses														
Sex-based HL PAFs[a]	42,654	3.1	2.3–3.8	1338	988–1641	NA	NA	NA	3353	1078	835–1330	NA	NA	NA
Removing countries[b]	40,781	2.6	1.7–3.3	1050	708–1359	647	100	52–150	3222	805	565–1064	145	145	91–145
Males & Females														
0 to 4	12,497	0.8	0.5–1.0	96	64–120	192	30	16–44	78	48	37–58	18	18	11–18
5 to 9	8504	2.6	1.8–3.2	218	157–270	238	37	19–55	262	160	125–194	21	21	13–21
10 to 14	8604	5.0	3.0–7.0	426	254–600	162	25	13–37	1189	355	212–517	46	46	29–46
15 to 19	13,062	2.7	2.0–3.3	357	267–429	88	14	8–21	1824	274	215–339	69	69	44–69
Males														
Overall: 0 to 19	22,453	3.0	2.0–3.9	676	456–867	521	81	43–120	1822	488	346–640	107	107	67–107
0 to 4	6680	1.1	0.7–1.4	73	50–92	144	22	12–33	65	40	31–48	11	11	7–11
5 to 9	4661	3.5	2.5–4.4	163	117–203	187	29	15–43	198	121	94–147	13	13	8–13
10 to 14	4424	5.4	3.2–7.5	239	142–333	122	19	10–28	623	186	111–271	34	34	21–34
15 to 19	6688	3.0	2.2–3.6	201	147–239	68	11	6–16	936	141	110–174	49	49	31–49
Females														
Overall: 0 to 19	20,214	2.1	1.4–2.7	421	286–552	159	25	13–37	1531	349	243–468	47	47	30–47
0 to 4	5817	0.4	0.2–0.5	23	14–28	48	8	4–11	13	8	6–10	7	7	4–7
5 to 9	3843	1.4	1.0–1.7	55	40–67	51	8	4–12	64	39	31–47	8	8	5–8
10 to 14	4180	4.5	2.7–6.4	187	112–267	40	6	3–9	566	169	101–246	12	12	8–12
15 to 19	6374	2.4	1.9–3.0	156	120–190	20	3	2–5	888	133	105–165	20	20	13–20

EBV = Epstein-Barr virus, CI = confidence interval, N = total number of estimated cases per age-sex group, PAF = population attributable fraction, AC = attributable cases due to the EBV, HL = Hodgkin lymphoma, NA = not applicable

a. Sex-based PAFs (41.7% for males and 20.8% for females) rather than age-based PAFs, were applied to cancer incidence.

b. Countries with less adequate cancer incidence data (Albania, Bosnia & Herzegovina, Denmark, Finland, Hungary, Latvia, Moldova, Montenegro, North Macedonia, Romania, and Serbia) were removed to assess if/how the overall PAF for EBV changes.

Supplementary material to manuscript #3

Acknowledgements

We are grateful to the following individuals who provided additional data

Dr. Caroline Besson (University Paris-Saclay, UVSQ, InsermU1018, CESP, Villejuif, France; Service d'hématologie et oncologie, Centre Hospitalier de Versailles, Le Chesnay – Rocquencourt, France); **Dr. Ellen Chang** (Center for Health Sciences, Exponent Inc., California, United States); **Dr. Laurence de Leval** (Institute of Pathology, Department of Laboratory Medicine and Pathology, Lausanne University Hospital and Lausanne University, Lausanne, Switzerland); **Dr. Arjan Diepstra** (Department of Pathology and Medical Biology, University Medical Center Groningen, the Netherlands); **Dr. Cecilie Dupont Harwood** (Department of Otorhinolaryngology, Head and Neck Surgery, and Audiology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark); **Dr. Peter Hollander** (Clinical and Experimental Pathology, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden); **Dr. Sam Mbulaiteye** (Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, United States); **Dr. Antonia Pavlovic** (Department of Pathology, Forensic Medicine and Cytology, Clinical Hospital Center, Split, Croatia); **Dr. Mounir Trimèche** (Institute of Pathology, Department of Laboratory Medicine and Pathology, Lausanne University Hospital and Lausanne University, Lausanne, Switzerland)

Appendix A. Electronic database search strategies performed on July 27, 2021

Embase Classic + Embase, 1947–	MEDLINE(R) 1946–
Population	
exp child/ OR exp pediatrics/ OR child$.mp. OR pediatric$.mp. OR paediatric$.mp. OR prematur*.mp. OR preterm*.mp. OR perinat$.mp. OR neonat$.mp. OR newborn$.mp. OR new born$.mp. OR infan$.mp. OR bab$.mp. OR toddler$.mp. OR boy$.mp. OR girl$.mp. OR kid$1.mp. OR school$.mp. OR juvenil$.mp. OR underage$.mp. OR under age$.mp. OR teen$.mp. OR minor$.mp. OR youth$.mp. OR pubescen$.mp. OR adolescen$.mp. OR infan$.jx. OR child$.jx. OR pediatric$.jx. OR paediatric$.jx. OR adolescen$.jx.	exp child/ OR exp pediatrics/ OR child$.mp. OR pediatric$.mp. OR paediatric$.mp. OR prematur*.mp. OR preterm*.mp. OR perinat$.mp. OR neonat$.mp. OR newborn$.mp. OR new born$.mp. OR infan$.mp. OR bab$.mp. OR toddler$.mp. OR boy$.mp. OR girl$.mp. OR kid$1.mp. OR school$.mp. OR juvenil$.mp. OR underage$.mp. OR under age$.mp. OR teen$.mp. OR minor$.mp. OR youth$.mp. OR pubescen$.mp. OR adolescen$.mp. OR infan$.jw. OR child$.jw. OR pediatric$.jw. OR paediatric$.jw. OR adolescen$.jw.
Exposure	
exp Herpesvirus 4, Human/ OR exp Epstein-Barr Virus Infections/ OR herpesvirus type 4.tw,kw. OR herpesvirus 4.tw,kw. OR ebv.tw,kw. OR ((epstein-Barr or epstein Barr) adj2 (virus* or viral*)).tw,kw. OR HHV4.tw,kw.	exp Herpesvirus 4, Human/ OR exp Epstein-Barr Virus Infections/ OR herpesvirus type 4.tw,kf. OR herpesvirus 4.tw,kf. OR ebv.tw,kf. OR ((epstein-Barr or epstein Barr) adj2 (virus* or viral*)).tw,kf. OR HHV4.tw,kf.
Outcome	
exp neoplasm/ OR (cancer* or neoplas* or tumor* or tumour* or malignan* or carcinoma* or metasta* or oncolog* or leukemi* or leukaemi* or lymphoma* or myeloma* or sarcoma* or squamous cell* or adenocarcinoma*).tw,kw.	exp Neoplasms/ OR (cancer* or neoplas* or tumor* or tumour* or malignan* or carcinoma* or metasta* or oncolog* or leukemi* or leukaemi* or lymphoma* or myeloma* or sarcoma* or squamous cell* or adenocarcinoma*).tw,kf.
Limits	
English AND humans	English AND humans

Appendix B: Abstracts reviewed for records

Conference name, years
AACR Annual Meeting, 2021, 2020, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011
AACR Special Conference on the Microbiome, Viruses, and Cancer, 2020
AACR Special Conference on the Advances in Pediatric Cancer Research, 2019
AACR Special Conference: Advances in Pediatric Cancer Research: From Mechanisms and Models to Treatment and Survivorship, 2015
American Society of Pediatric Hematology / Oncology 2021, 2020, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011
Biannual International Symposium on Nasopharyngeal Carcinoma, 2015
International Symposium on Childhood, Adolescent and Young Adult Non-Hodgkin Lymphoma, 2018, 2015, 2012, 2009
International Symposium on Hodgkin lymphoma, 2018, 2016, 2013, 2010
International Society of Paedatric Oncology, 2020, 2019, 2018, 2017, 2016, 2015

AACR = American Association for Cancer Research

Appendix C: Description of quality assessment

Here, we describe how each of the nine items in the Joanna Briggs Institute (JBI) prevalence critical appraisal tool were *or were not* applied to our study.[383] The JBI tool stipulates four possible answers: yes, no, unclear, not applicable – we also included 'not reported' as an option. To receive a combined quality score, each 'yes' response was awarded a point.

As written in the JBI tool	How it was applied to our study *(options)*	What qualified as appropriate/adequate/yes
1. Was the sample frame appropriate to address the target population?	Setting where patients were recruited from: *single study site/location, multiple study sites/locations, population-based cancer registry/ies*	• Multiple study sites/locations • Population-based registry/ies
2. Were study participants sampled in an appropriate way?	How cases were enrolled/sampled: *consecutive/complete, random, convenience, not reported*	• Consecutive/complete • Random
3. Was the sample size adequate?	Number of patients aged 0 to 19 with valid EBV testing results	• 20 for Hodgkin lymphoma • 8 for all other cancer sites/types
4. Were the study subjects and the setting described in detail?	Reported the sex of patients When specimen collection took place in relation to treatment: *pretreatment, biopsy for diagnostic reasons, not reported*	One point each for • Sex reported for the relevant part of the study population[a] • Reported that specimens collected pre-treatment or indicating it was a biopsy taken for diagnosis
5. Was the data analysis conducted with sufficient coverage of the identified sample?	Recorded the number of samples available for testing	• Reported how many specimens were available from the identified sample AND that <20% of the specimens were not available for testing and with valid testing
6. Were valid methods used for the identification of the condition?	Recorded the method(s) used to detect EBV	• Applied as an inclusion criterion
7. Was the condition measured in a standard, reliable way for all participants?	Recorded information on how EBV was measured	• The method of testing for EBV was the same for all patients
8. Was there appropriate statistical analysis?	Not applicable as we performed the statistical analysis (i.e., calculation of confidence intervals)	• Not relevant • To be included, the numerator (# testing EBV+) and denominator (# tested for EBV) had to be identifiable (or study authors provided them upon request)
9. Was the response rate adequate, and if not, was the low response rate managed appropriately?	Not applicable as the source of cases is often pathology records	• Not relevant

JBI = Joanna Briggs Institute, EBV = Epstein-Barr virus

a. Twenty-five of the 49 included studies (51.0%) reporting on BL, HL or NPC had study populations that included adults in addition to children/adolescents.

Table S1. Percentage of all cancers diagnosed among individuals aged 0 to 19 attributable to EBV, by country

Method

1. GLOBOCAN incidence data with Burkitt lymphoma (BL) proportion derived IICC data from the same country.
2. GLOBOCAN incidence data with BL proportion derived from IICC data from other countries (countries used to estimate the BL proportion).
3. No non-Hodgkin lymphoma cases to partition.
4. No GLOBOCAN data available and therefore excluded from the analysis.

Country	Method	Country/ies BL proportion derived from	Total N	Attributable n	Percent of all cancers attributable to EBV %
Europe					
Albania	2	Croatia, Greece, Italy	91	<1	0.5
Andorra	4		--	--	--
Austria	1		295	2	0.8
Belarus	1		323	8	2.3
Belgium	1		573	16	2.8
Bosnia & Herzegovina	2	Croatia, Greece, Italy	44	2	4.6
Bulgaria	1		201	6	3.1
Croatia	1		159	2	1.5
Cyprus	3		41	2	4.1
Czechia	1		295	3	1.1
Denmark	2	Norway & Sweden	230	5	2.2
Estonia	1		34	<1	0.6
Finland	2	Norway & Sweden	185	3	1.8
France	1		2840	88	3.1
Germany	1		2922	71	2.4
Greece	1	Italy (ages 15-19 only)	345	10	3.0
Greenland	4		--	--	--
Hungary	1	Austria (ages 15-19 only)	335	7	2.1
Iceland	3		7	0	0.0
Ireland	1		215	5	2.4
Italy	1		2333	75	3.2
Kosovo	4		--	--	--
Latvia	1		58	2	3.3
Lithuania	1		78	1	1.4
Liechtenstein	4		--	--	--
Luxembourg	3		12	<1	1.3
Malta	3		7	<1	2.1
Moldova (Republic of)	2	Hungary & Ukraine	102	<1	1.2
Monaco	4		--	--	--
Montenegro	2	Croatia, Greece & Italy	23	<1	1.3
Netherlands	1		622	15	2.3
North Macedonia	2	Greece & Bulgaria	69	2	3.4
Norway	1		204	5	2.2
Poland	1		1190	23	1.9
Portugal	1		344	15	4.4
Romania	2	Bulgaria, Hungary & Ukraine	393	16	4.0
Russian (Federation)	1		5019	134	2.7
San Marino	4		--	--	--
Serbia	2	Croatia, Bulgaria & Hungary	359	7	2.0
Slovakia	1		213	6	2.7
Slovenia	1		53	1	2.2
Spain	1		1433	49	3.4
Sweden	1		332	6	1.8
Switzerland	1		302	8	2.6
Ukraine	1		1342	47	3.5
United Kingdom	1		2840	72	2.5
North America					
Canada	1		1496	33	2.2
United States	1		14,691	343	2.3

Appendix D: Information on individual studies

Table S2. Studies reporting on EBV prevalence as detected by EBER ISH in BL tumor tissues collected from individuals aged 0 to 19 residing in Europe or North America

Study	Country/ies	Specimen collection dates	Quality score[a]	Male %	Age range	Cases N	EBV+ %
Richter 2022[296]	Germany	2001–2013	5	86.8	≤18	89	5.6
Dupont 2021[297]	Denmark	1980–2018	5	81.8	3–19	22	13.6
Mbulaiteye 2013[140]	US	1979–2009	6	91.3	0–19	23	13.0
Kasprzak 2007[298]	Poland	1999–2003	5	92.9	3–16	14	57.1
Karajannis 2005[299]	Austria, Germany & Switzerland	1990–1998	6	79.7	1–18	222	11.3
Teitell 2005[300]	France & United Kingdom	NS	5	85.7	2–16	14	28.6
Haralambieva 2004[301]	the Netherlands	NS	4	NS	5–13	13	23.1

EBV = Epstein-Barr virus, BL = Burkitt lymphoma, EBER ISH = Epstein-Barr encoding region *in situ* hybridization, NS = Not specified, US = United States

a. Quality scored based on seven items (a study could be awarded a maximum of seven points); the quality assessment is described in appendix C.

Table S3. Studies reporting on EBV prevalence in HL tumor tissues collected from individuals aged 0 to 19 residing in Europe or North America

Study (reference)	Country/ies	Specimen collection years	Male, %	Age range	Detection method(s)	Quality score (out of 7)[a]	Cases N	Positive %	PAF estimates	
									Main	Sex-based
Hamdi 2021[401]	France	2008–2010	52.6	4–18	EBER and/or LMP-1	6	57	21.1	x	x
Dilly-Feldis 2019[396]	France	1997–2014	NS	4–18	EBER and/or LMP-1	5	41	26.8		
Bigenwald 2017[392]	France	1979–2013	NS	15–20	LMP-1	3	78	10.3	x	
Hollander 2017[404]	Denmark & Sweden	1990–2007	53.3	15–19	EBER and/or LMP-1	5	30	23.3	x	x
Englund 2016[398]	Sweden	1983–2008	50.6	3–17	EBER and/or LMP-1	7	87	25.3	x	x
Pavlovic 2016[415]	Croatia	1997–2009	35.7	5–19	LMP-1	3	14	21.4	x	x
Linabery 2015[312]	Canada & US	1989–2003	NS	0–14	EBER	3	355	23.7	x	
Huppmann 2014[406]	Canada & US	2000–2013	91.7	0–18	EBER and/or LMP-1	5	145	3.4		
Klekawka 2013[411]	Poland	NS	NS	3–18	EBER and/or LMP-1	3	61	44.3		
Horton 2012[405]	US	2003–2006	NS	1–20	EBER and/or LMP-1	6	102	25.5		
Siddon 2012[313]	US	NS	50.0	12–17	EBER	2	10	0.0		
Trimèche 2009[417]	Belgium	1989–2004	60.0	8–19	EBER	4	15	26.7		x
Glaser 2008[314]	US	1988–1997	NS	0–19	EBER and/or LMP-1	4	131	35.1	x	
Heller 2008[315]	US	NS	45.5	7–19	EBER	5	22	54.5	x	x
Diepstra 2007[395]	Netherlands	1989–2000	54.3	NS	EBER	6	45	17.8	x	x
Lacroix 2007[413]	France	NS	NS	8–18	LMP-1	1	12	41.7		
Claviez 2005[393]	Germany	1990–2001	55.2	2–20	LMP-1	4	842	31.2	x	x
Keegan 2005[146]	US	1988–1997	61.1	0–14	EBER and/or LMP-1	5	36	44.4		x
Chang 2004[147]	US	1997–2001	40.0	15–19	EBER and/or LMP-1	4	10	20.0	x	
Herling 2003[403]	Greece, Italy & US	1984–2000	NS	0–16	LMP-1	5	27	29.6		
Jarrett 2003[407]	UK (Scotland)	1993–1997	NS	3–18	EBER and/or LMP-1	6	40	20.0	x	
Flavell 2001[399]	UK (England)	1981–1999	67.3	0–14	LMP-1	5	55	61.8	x	x

EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, PAF = population attributable fraction, NR = not specified, EBER ISH = Epstein-Barr encoding region *in situ* hybridization; LMP-1 = latent membrane protein 1, US = United States, UK = United Kingdom

[a.] Quality scored based on seven items (a study could be awarded a maximum of seven points); the quality assessment is described in appendix C.

Table S3. Studies reporting on EBV prevalence in HL tumor tissues collected from individuals aged 0 to 19 residing in Europe or North America (continued)

Study	Country/ies	Specimen collection years	Male, %	Age range	Detection method(s)	Quality score (out of 7)[a]	Cases N	Positive %
Enblad 1999[397]	Sweden	1985–1988	NS	11–19	LMP-1	3	15	80.0
Armstrong 1998[391]	UK (Scotland)	NS	NS	0–19	EBER and/or LMP-1	4	73	20.5
Santon 1998[416]	Spain	NS	70.8	3–18	LMP-1	5	24	100.0
Andriko 1997[316]	US	1984–1996	90.9	3–15	LMP-1	5	44	38.6
Razzouk 1997[317]	US	NS	42.3	5–18	EBER	4	26	57.7
Herbst 1996[402]	Germany	NS	NS	0–20	LMP-1	3	49	34.7
Kordek 1996[412]	Poland	1986–1993	NS	0–14	LMP-1	3	10	60.0
Lin 1996[318]	US	1971–1992	NS	3–18	EBER	3	10	30.0
Panayiotides 1996[414]	Greece	1984–1987	NS	3–18	LMP-1	2	13	100.0
Weinreb 1996[419]	Greece	1972–1991	NS	2–14	LMP-1	2	22	90.9
Claviez 1994[394]	Germany	1976–1992	57.1	4–17	EBER and/or LMP-1	5	21	47.6
Kaczorowski 1994[408]	Poland	NS	NS	0–14	LMP-1	5	32	53.1
Kanavaros 1994[409]	Greece	1984–1987	NS	3–18	EBER and/or LMP-1	3	22	54.5
Ambinder 1993[390]	US	NS	48.0	NS–15	EBER and/or LMP-1	6	25	36.0
Brousset 1993[387]	France	NS	NS	8–15	EBER and/or LMP-1	1	13	53.8
Foss 1993[400]	Germany & Italy	NS	75.0	8–19	EBER	5	8	25.0
Khan 1993[410]	UK (England)	NS	NS	0–14	EBER and/or LMP-1	3	24	25.0
Weinreb 1992[418]	UK	1957–1992	73.0	0–15	LMP-1	5	74	50.0

EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, CI = confidence interval, EBER ISH = Epstein-Barr encoding region *in situ* hybridization; LMP-1 = latent membrane protein 1, NS = not specified, PAF = population attributable fraction, US = United States, UK = United Kingdom

a. Quality scored based on seven items (a study could be awarded a maximum of seven points); the quality assessment is described in appendix C.

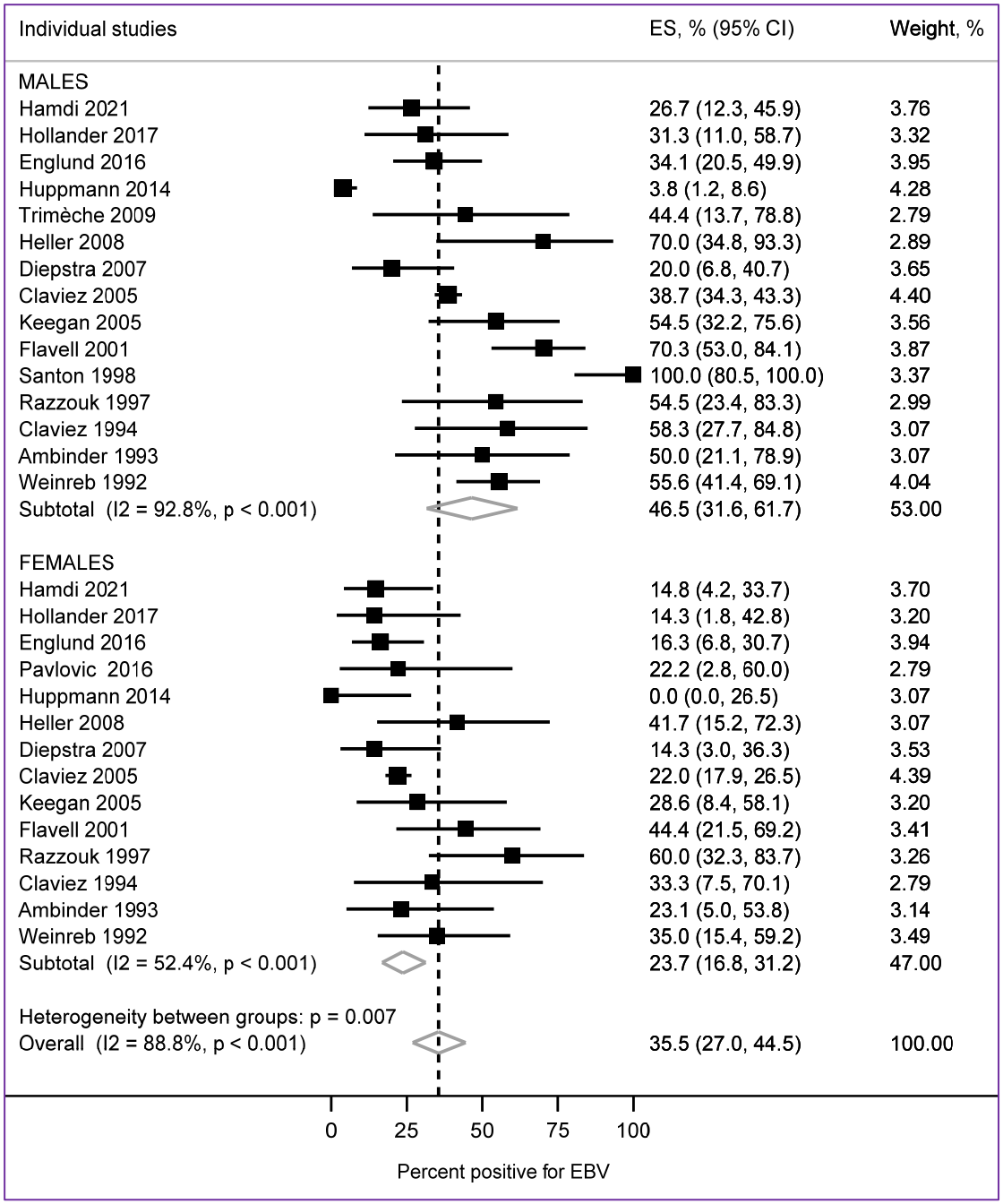
Table S4. Studies reporting on EBV prevalence as detected by EBER ISH in NPC tumor tissues collected from individuals aged 0 to 19 residing in Europe or North America[a]

Study	Country	Specimen collection dates	Quality score[a]	Male %	Age range	Cases N	Positive %
Polychronopoulou 2004[320]	Greece	1987–2001	5	NS	7–14	8	100.0
Mertens 1997[319]	Germany	1992–NR	3	NS	NS	8	100.0

EBV = Epstein-Barr virus, NPC = nasopharyngeal carcinoma, EBER ISH = Epstein-Barr encoding region *in situ* hybridization, NR = Not specified

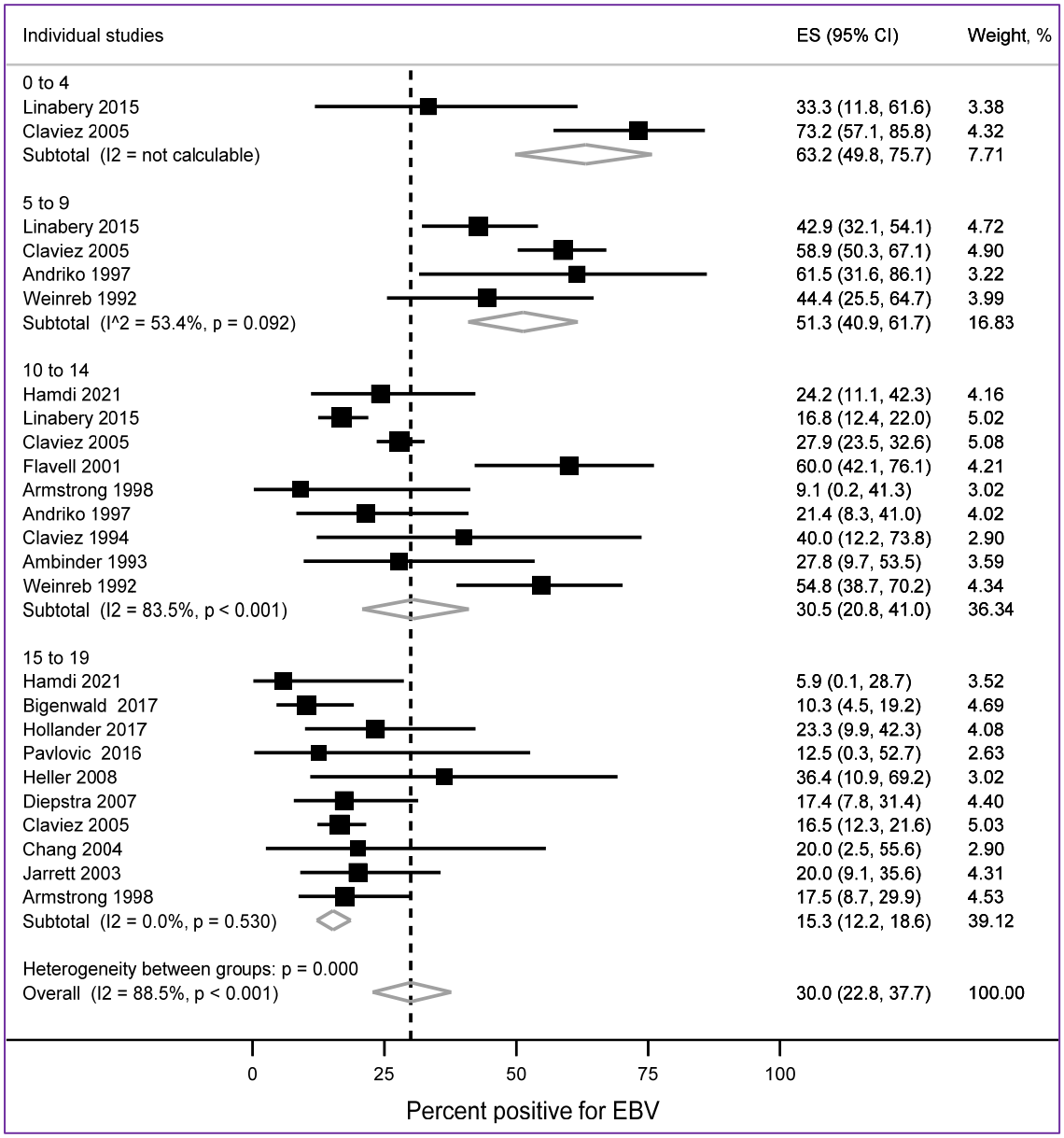
[a.] Quality scored based on seven items (a study could be awarded a maximum of seven points); the quality assessment is described in appendix C.

Fig. S1. Forest plot of EBV prevalence (%) in HL tumor tissues collected from individuals aged 0–19 residing in Europe or North America, by sex



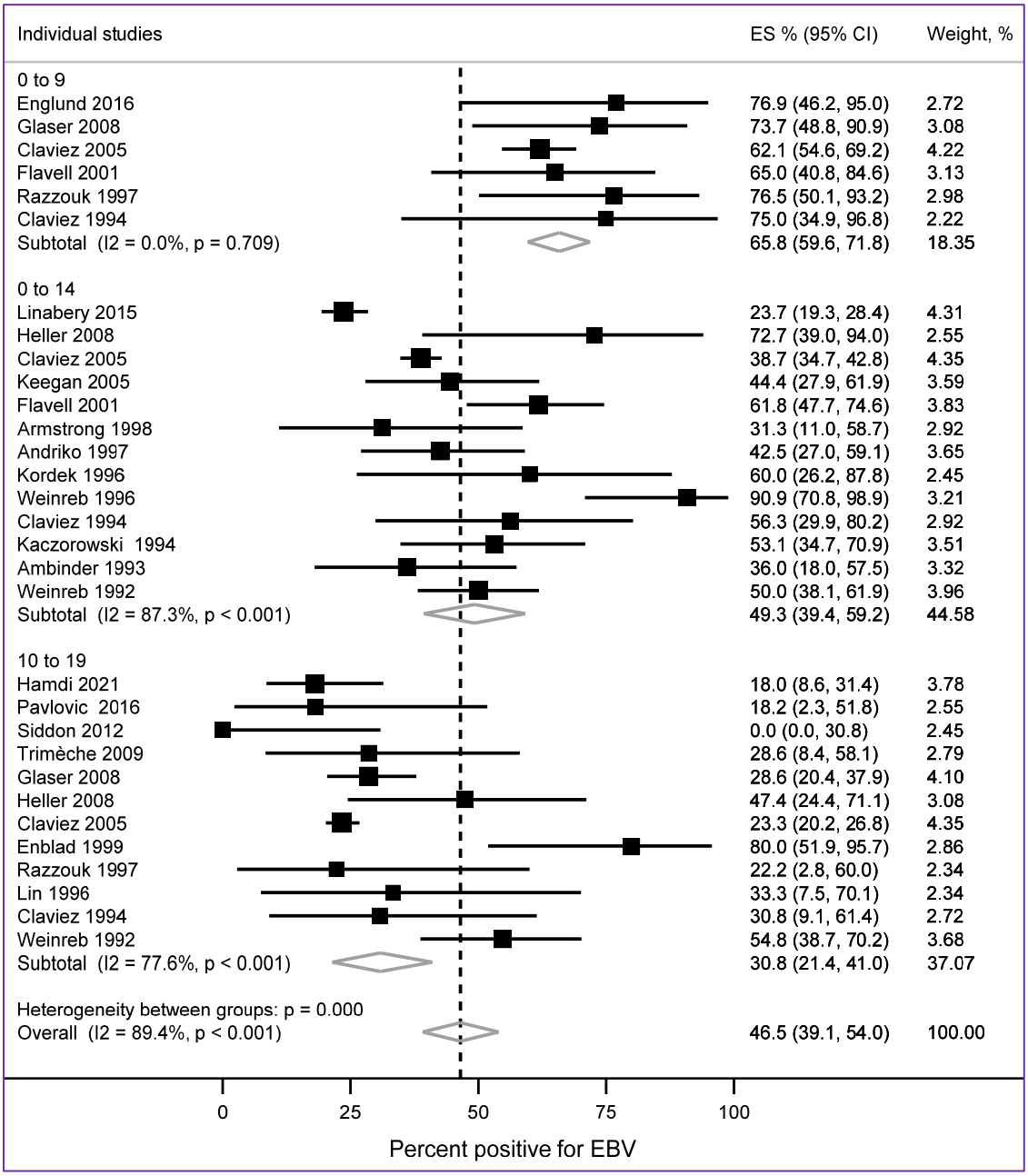
EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, CI = confidence interval

Fig. S2. Forest plot of EBV prevalence (%) in HL tumor tissues collected from individuals aged 0–19 residing in Europe or North America, by 5-year age groups



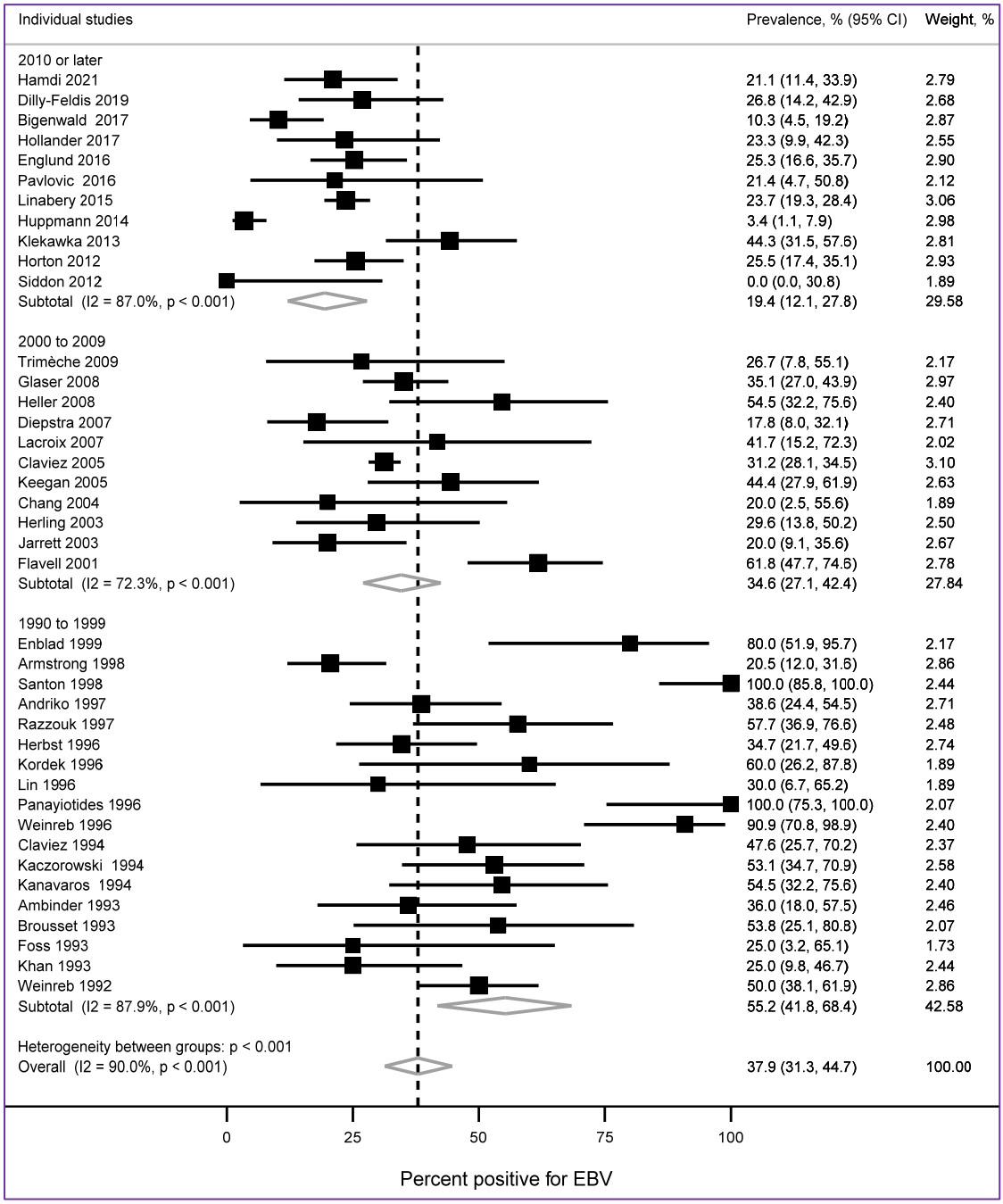
EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, CI = confidence interval

Fig. S3. Forest plot of EBV prevalence (%) in HL tumor tissues collected from individuals aged 0–19 residing in Europe or North America, by 10 or 15-year age groups



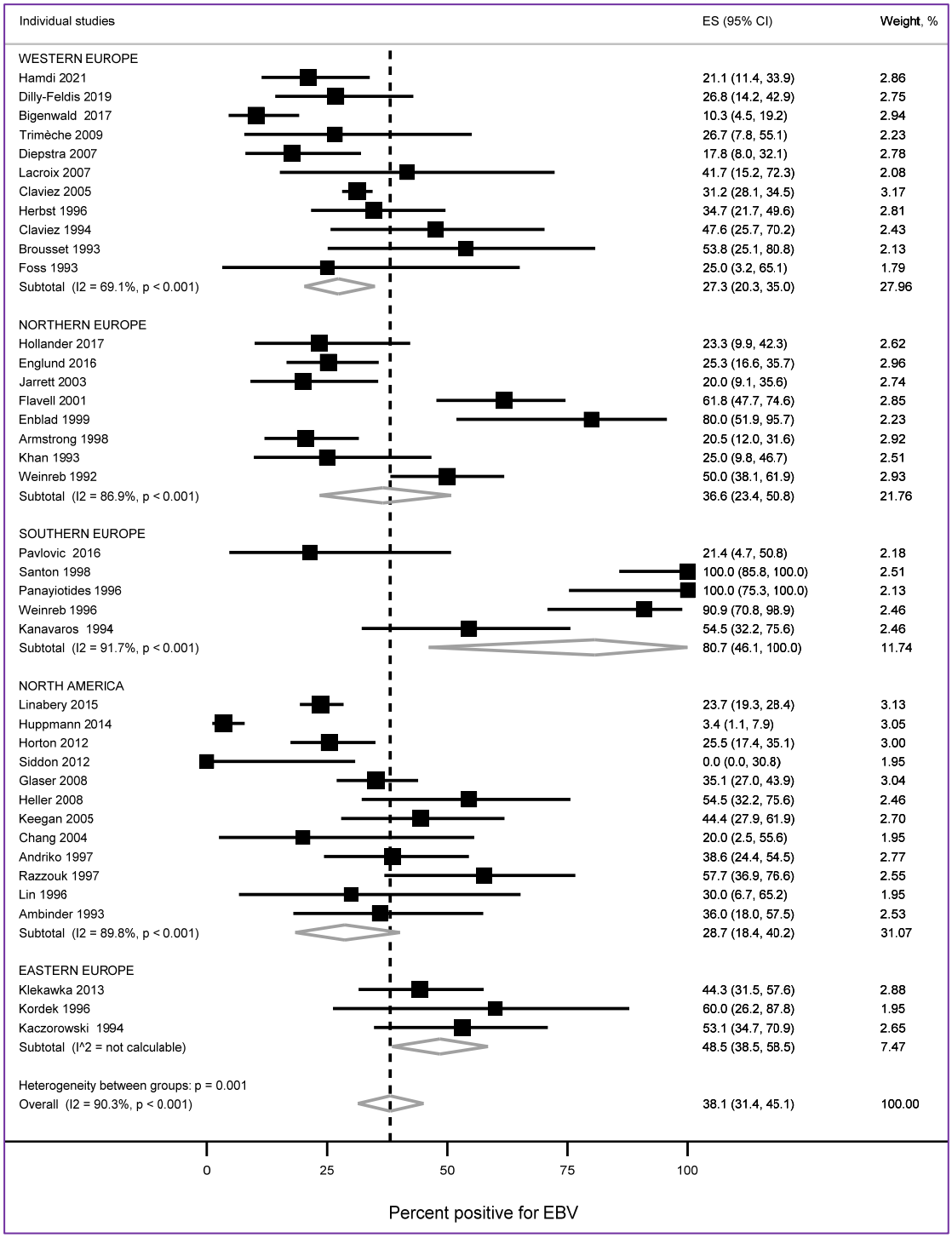
EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, CI = confidence interval

Fig. S4. Forest plot of EBV prevalence (%) in HL tumor tissues collected from individuals aged 0–19 residing in Europe or North America, by year of publication



EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, CI = confidence interval

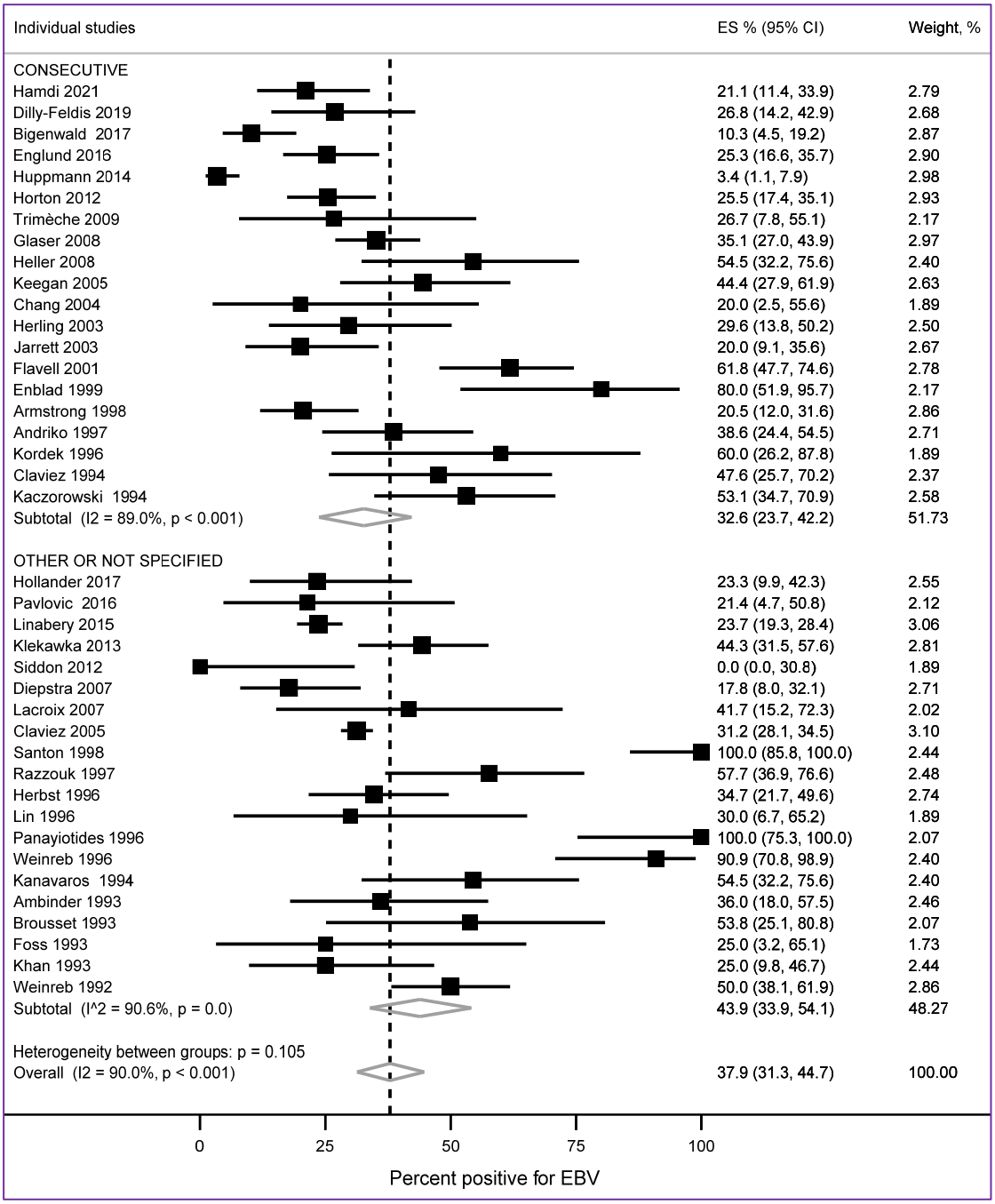
Fig. S5. Forest plot of EBV prevalence (%) in HL tumor tissues collected from individuals aged 0–19 residing in Europe or North America, by region[a]



EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, CI = confidence interval

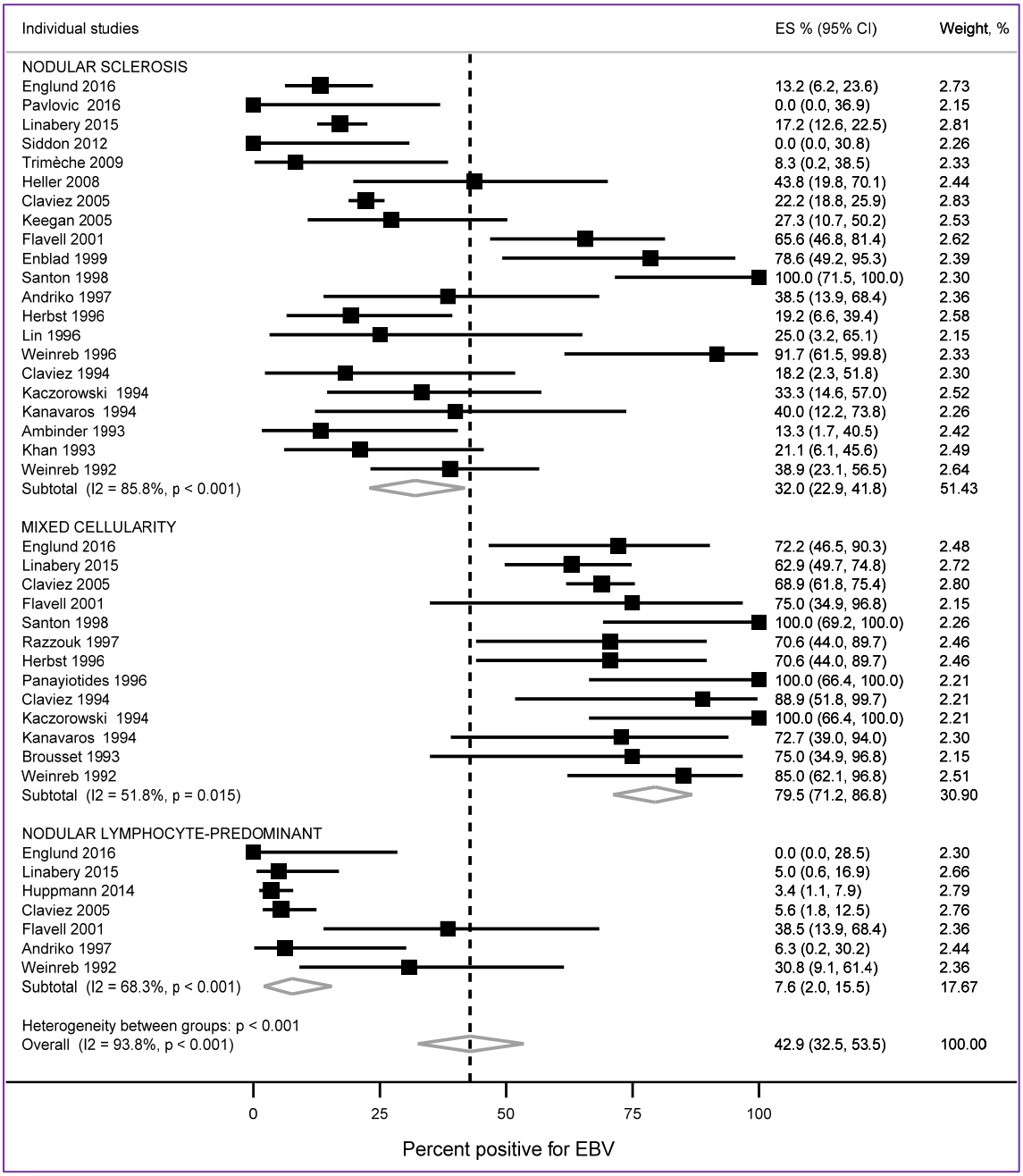
a. One study was conducted in both continents (Greece, Italy, & US) and was excluded from this forest plot.[403]

Fig. S6. Forest plot of EBV prevalence (%) in HL tumor tissues collected from individuals aged 0–19 residing in Europe or North America, by method used to enroll cases



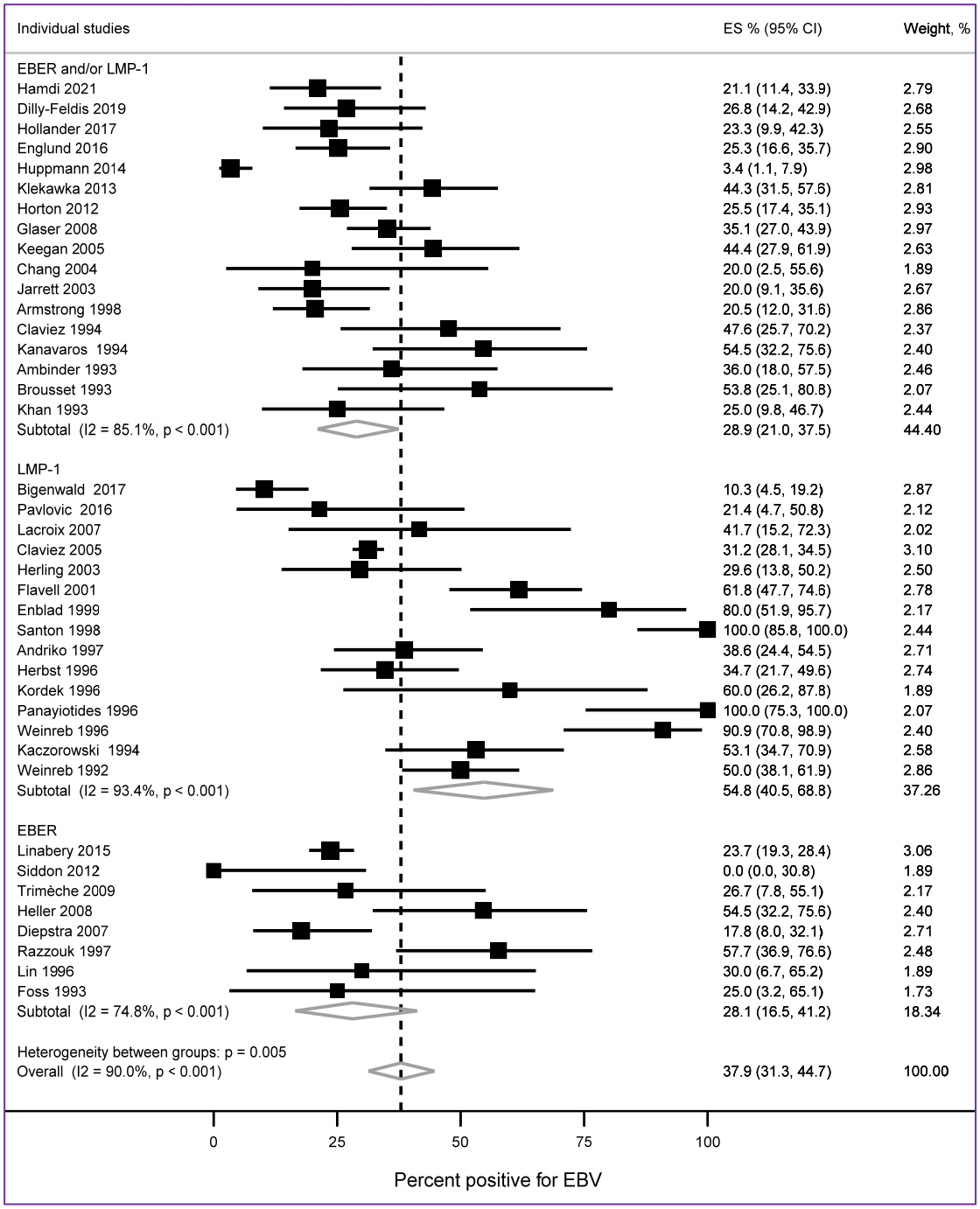
EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, CI = confidence interval

Fig. S7. Forest plot of EBV prevalence (%) in HL tumour tissues collected from individuals aged 0–19 residing in Europe or North America, by HL subtype



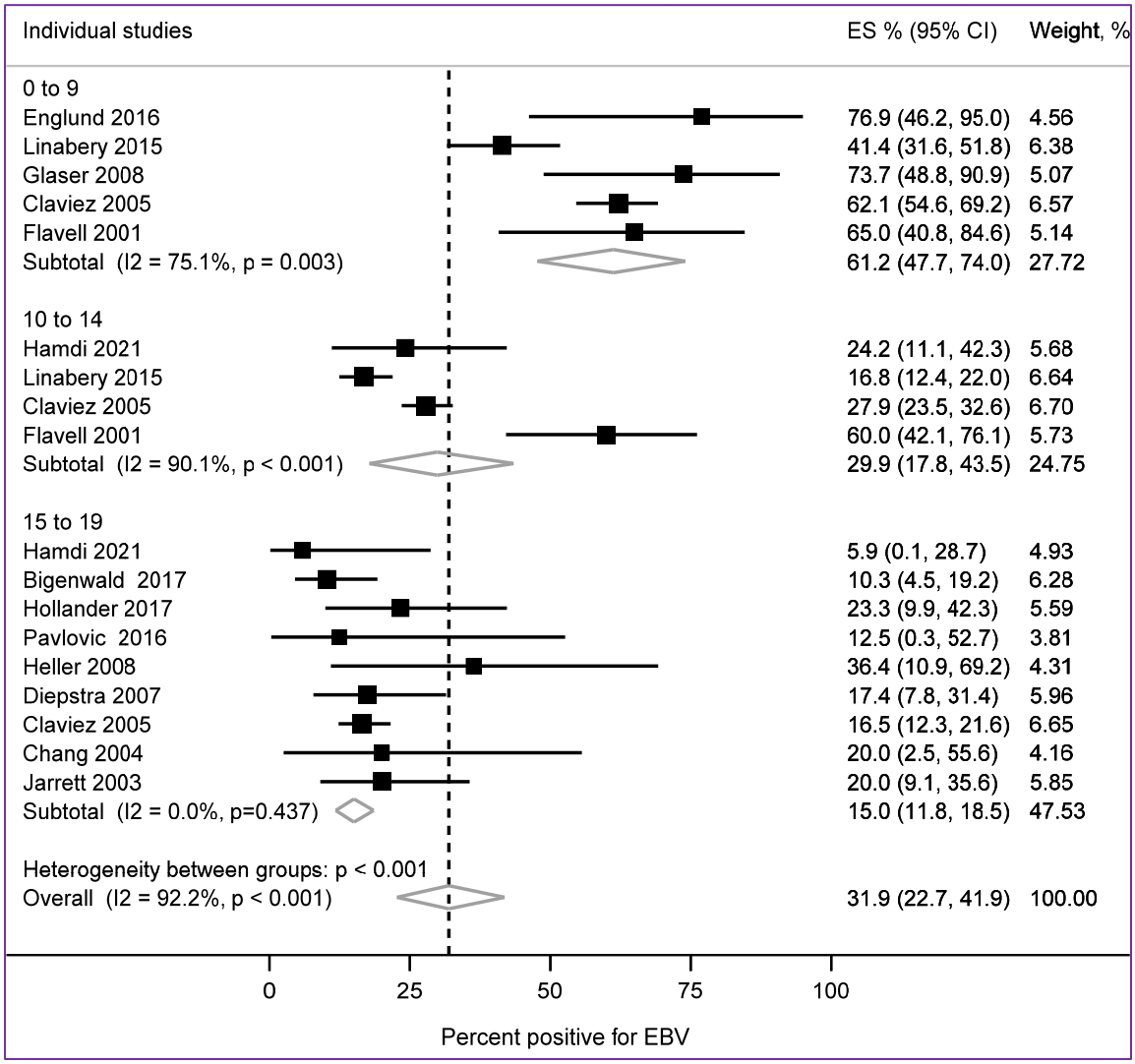
EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, CI = confidence interval

Fig. S8. Forest plot of EBV prevalence in HL tumor tissues collected from individuals aged 0–19 residing in Europe or North America, detection method



EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, CI = confidence interval, EBER ISH = Epstein-Barr encoding region *in situ* hybridization, LMP-1 = latent membrane protein 1

Fig. S9. Forest plot of EBV prevalence in HL tumor tissues collected from individuals aged 0–19 residing in Europe or North America, age-groups used in PAF analyses



EBV = Epstein-Barr virus, HL = Hodgkin lymphoma, CI = confidence interval, PAF = population attributable fraction

Chapter 4: Future Burden of Infection Attributable Cancers in Canada

This study was situated in the ComPARe Study and contributed to its second goal: to quantify the annual number of incident cancer cases that would occur between 2018 and 2042 and the potential impact of prevention initiatives on that cancer incidence.

Manuscript #4: Estimates of the future burden of cancer attributable to infections in Canada

This manuscript includes estimates of the impact that various counterfactual reductions in infection prevalence (HBV, HCV, and *H. pylori*) and levels of school-based HPV vaccine uptake could have on cancer incidence in Canada up to and including the year 2042. The manuscript was published as part of a special issue on the burden of cancer in Canada, in the journal Preventive Medicine in the spring of 2019. The published version of this manuscript can be found in the appendix.

Estimates of the future burden of cancer attributable to infections in Canada

Karena D. Volesky[1,2]
Mariam El-Zein[1]
Eduardo L. Franco[1,2]
Darren R. Brenner[3]
Christine M. Friedenreich[4]
Yibing Ruan[4]
On behalf of the ComPARe Study Team[5]

1. Gerald Bronfman Department of Oncology, Division of Cancer Epidemiology, McGill University, Montréal, Québec, Canada
2. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada
3. Departments of Oncology and Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
4. Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada
5. Additional ComPARe Study Team members: **Stephen Walter**, Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada; **Will King**, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; **Paul Demers**, Occupational Cancer Research Centre, Toronto, Ontario, Canada; **Paul Villeneuve**, Department of Health Sciences, Carleton University, Ottawa, Ontario, Canada; **Abbey Poirier**, Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada; **Prithwish De**, Cancer Care Ontario, Toronto, Ontario, Canada; **Leah Smith**, Canadian Cancer Society, Toronto, Ontario, Canada; Elizabeth Holmes, Cancer Care Ontario, Toronto, Ontario, Canada; **Dylan O'Sullivan**, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; **Zeinab El-Masri**, Cancer Care Ontario, Toronto, Ontario, Canada; **Robert Nuttall**, Canadian Cancer Society, Toronto, Ontario, Canada; **Sheila Bouten**, Division of Cancer Epidemiology, McGill University, Montréal, Québec, Canada; **Elizabeth Holmes**, Canadian Cancer Society, Toronto, Ontario, Canada; **Tasha Narain**, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; **Priyanka Gogna**, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada

This manuscript is presented in the format it is published in the journal, Preventive Medicine, with some minor modifications to appear more consistent with the other manuscripts contained in this thesis.

Volesky KD, El-Zein M, Franco EL, Brenner DR, Friedenreich CM, Ruan Y. Estimates of the future burden of cancer attributable to infections in Canada. Prev Med. 2019;122:118-27

Acknowledgements: Karena D. Volesky is supported by a doctoral award from the Fonds de recherche du Québec – Santé (FRQS) and Fondation des étoiles. Darren Brenner holds a Canadian Cancer Society Prevention Capacity Development Award (#703917). Christine Friedenreich was supported by an Alberta Innovates Health Senior Scholar Award and by the Alberta Cancer Foundation Weekend to End Women's Cancers Breast Cancer Chair. We are grateful to Talía Malagón for her modelling advice, and to Shemay Lee for assisting us with data management.

Disclosure: E.L.F. has served as occasional consultant to companies involved with HPV diagnostics and vaccination (Merck, GSK, Roche, and BD). His institution has received grants from Merck and Roche to supplement investigator-initiated studies that he leads at McGill University.
E.L.F. is Editor-in-Chief at Preventive Medicine and K.D.V. is an Assistant Editor at Preventive Medicine. The process of soliciting the special issue, sending out manuscripts for review, the peer-review process and editorial decision making was conducted entirely outside of the Preventive Medicine online system (for which E.L.F. and K.D.V. have access to through their regular Preventive Medicine duties).

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

Role of funding source: The sponsor had no role in the study design, data collection, data analysis or interpretation, writing of the manuscript or in the decision to submit this manuscript for publication.

HIGHLIGHTS

- The major cancer-causing infections can be prevented or treated.
- Reducing hepatitis C virus prevalence by 50% can prevent 1190 cancers by 2042.
- Reducing *H. pylori* prevalence by 50% can prevent 8,700 cancers by 2042.
- Over 5000 cancers could be prevented by 2042 with 80% HPV vaccine coverage.

ABSTRACT

More than 7,000 incident cancers diagnosed in Canada in 2015 were attributable to infections. The future infection-associated cancer burden can be lowered by reducing the prevalence of major cancer-causing infections, hepatitis B virus (HBV), hepatitis C virus (HCV), *Helicobacter pylori* (*H. pylori*) and human papillomavirus (HPV). We modeled the future impact of (1) 10%, 25%, and 50% relative reductions in the prevalence of HBV, HCV and *H. pylori* and (2) different levels (lower, current, higher) of school-based HPV vaccination coverage on Canadian cancer incidence by the year 2042. We modeled counterfactual reductions in HBV, HCV and *H. pylori* prevalence in 2018 assuming a latency period of 15-years to estimate impacts on cancer incidence starting in 2033. The number of HPV-attributable cancers among vaccinated cohorts was a function of pre-2018 vaccine coverage levels and the 2018 counterfactuals. A 50% counterfactual reduction in the prevalence of HBV, HCV and *H. pylori* could prevent an estimated 10,585 cancers from 2018 to 2042; a 25% reduction could prevent 5,293 cancers and a 10% reduction could prevent 2,117 cancers. Assuming continuity of current estimated country-wide HPV vaccine coverage, 3,977 anogenital and 1,073 head and neck cancers could be prevented from 2018 to 2042, whereas vaccine coverage of 80% in girls and boys could prevent an additional 310 cancers. Almost 16,000 cancers could be prevented in Canada from 2018 to 2042 with a 50% relative reduction in HBV, HCV and *H. pylori* prevalence and 80% HPV vaccine coverage of girls and boys.

1. INTRODUCTION

Globally, an estimated 14.0% of cancers diagnosed in 2012 were attributable to four infectious agents; hepatitis B virus (HBV), hepatitis C virus (HCV), *Helicobacter pylori* (*H. pylori*) and human papillomavirus (HPV).[56] Several strategies have been adopted to reduce the prevalence of cancer-causing infections and their associated cancer or pre-cancer incidence in Canada and abroad. Canadian provinces/territories introduced publicly-funded, school-based immunization programs for HBV from 1992 to 1998 and for HPV from 2007 to 2010.[8,9] Due to HBV's long latency, reductions in cancer incidence have not yet been realized. However, the annual number of reported HBV infections in Canada has decreased from 10.8 per 100,000 persons in 1990 to 1.7 per 100,000 persons in 2008.[428] A meta-analysis of 20 ecologic population-based studies conducted in high-income countries reported a 68% decrease in the prevalence of HPV types 16 and 18 at a vaccination coverage among girls of 50% or higher.[429] A meta-analysis of randomized controlled trials reported that eradication of *H. pylori* in asymptomatic populations reduced gastric cancer risk by 34%.[430]

Despite infections' impact on global cancer incidence, the level of public awareness of a causal role for infections in the development of cancer is low. Yet, the public plays a key role by vaccinating their children against HBV and HPV, not reusing needles and complying with antibiotic treatment for *H. pylori* infection. The range of primary prevention strategies aimed at reducing the acquisition of infections (HBV, HCV and HPV) and secondary prevention strategies for treating existing infections (HCV, *H. pylori*) provides an opportunity to lower the infection-associated cancer burden.

We estimated that more than 7,000 cases of cancers, representing 3.7% of all cancers diagnosed among Canadians aged 18 and older in 2015 were attributable to seven carcinogenic infections.[431] The vast majority (90.0%) of these infection-attributable cancers were due to HBV, HCV, *H. pylori* and HPV. We found that, with ~3800 attributable cases, more cancers were attributed to HPV than any other infection. The infection with the next highest number of attributable cases was *H. pylori* with 2050 cases, followed by Epstein-Barr virus with 580 cases, hepatitis B and C virus with 510 cases, human herpesvirus type 8 (i.e., Kaposi sarcoma virus) with 100 cases and finally human T-cell lymphotropic virus type 1 with 30 attributable cases in 2015.

The considerable potential to prevent carcinogenic infections highlights the importance of quantifying the impact of a variety of prevention scenarios, referred to as counterfactuals, for prioritizing strategies aimed at reducing the number of infection-associated cancers. To our knowledge, besides the impact of HPV on cancer incidence,[432] no study has estimated the impact of reductions in the prevalence of infections on the future Canadian cancer incidence. We estimated the future burden of infection-associated cancers by the year 2042 by modelling the impact of: 1) relative reductions in HBV, HCV and *H. pylori* infection prevalence and 2) lower, current, and higher levels of school-based HPV vaccination coverage.

2. METHODS

This analysis is part of the Canadian population attributable risk of cancer (ComPARe) project, which aimed to estimate the current and future burden of cancer attributable to modifiable risk factors in Canada.[70] Here, we estimated the future burden of cancers caused by four major infectious agents (HBV, HCV, *H. pylori* and HPV). The future burden and the potential for the prevention of infection-associated cancers are reported as: the number of cancers projected and prevented in 2042 and the cumulative number of cancers prevented from 2018 to 2042 based on different counterfactuals.

We calculated potential impact fractions (PIFs) to estimate the proportion of HBV, HCV and *H. pylori*-associated incident cancers that could be avoided by 2042 under various counterfactual scenarios, using the following equation:[76]

$$PIF = \frac{(\mathrm{P} - \mathrm{P}^*)(\mathrm{RR} - 1)}{\mathrm{P}(\mathrm{RR} - 1) + 1}$$

where P is the pre-counterfactual infection prevalence, P* is the post-counterfactual infection prevalence, and RR is the relative risk or odds ratio (OR) between the infection and cancer. The annual prevented cases were estimated as:

$$PC_i = I_i \times PIF$$

where I_i is the projected cancer incidence in year *i*.

For HPV, we approximated the proportion of cancers attributable to HPV by using prevalence of HPV in cancer cases and therefore did not calculate PIFs. Knowing the proportion

of specific cancers attributable to HPV enabled us to estimate the number of avoidable HPV-related cancer cases. When estimating the future number of preventable HPV-associated cancers among vaccinated cohorts, the proportion attributable to HPV was subtracted, after accounting for vaccine efficacy, protection (e.g., the proportion of HPV types contributing to cancer incidence that are covered by the vaccines), and coverage.

2.1. Current infection prevalence

We have also reported on the prevalence of chronic HBV and HCV, and *H. pylori* for the Canadian population.[431] Briefly, chronic HBV prevalence (measured by hepatitis B surface antigen (HBsAg)), was assessed using data from two merged cycles (2007–2009 and 2009–2011) of the Canadian Health Measures Survey (CHMS).[92,97] Since we were only able to obtain sex-specific prevalence estimates from the CHMS, HBsAg prevalence from two merged cycles of the United States' National Health and Nutrition Examination Survey (NHANES) were used to partition the HBsAg sex prevalence estimates from the CHMS by 10-year age groups.[94,95] To estimate chronic HCV prevalence, we partitioned the five-year birth cohort estimates from a modelling study[90] according to the sex distribution reported in a study that modeled acute and chronic HCV prevalence in the Canadian population.[90,99] Since we required that prevalence estimates originate from population-based data covering a range of ages, the few studies assessing *H. pylori* sero-status in Canadian populations did not meet this criterion.[100-102] Hence, to estimate the prevalence of *H. pylori*, we reweighted NHANES data collected from 1999 to 2000[103] to reflect the Canadian age, sex, and race/ethnic composition (categories available were: Black, Latin American, White, and Other). To produce summary prevalence estimates, we calculated population-weighted prevalence estimates by sex thereby aggregating prevalence across age-groups (**Table 1**).

Rather than estimating HPV prevalence among the Canadian population, we estimated HPV prevalence among cancer cases. Since mechanistic evidence indicates that the detection of HPV within cancer tissue is sufficient to attributable that cancer to HPV, the population attributable risk (PAR) is approximated by the prevalence in cases. The prevalence of HPV infection was calculated by pooling, using a random effects model, the proportion of cancer cases harboring high-risk HPV types (for anogenital cancers) or HPV16 (head and neck cancers) within

the cancer tumor tissue. We restricted our analyses to studies that applied "gold standard" HPV detection techniques: polymerase chain reaction (PCR) for anogenital cancers and detection of E6 and/or E7 oncoproteins via PCR for head and neck cancers.[177,178]

Table 1 summarizes the prevalence of these infections in the population (for HBV, HCV and *H. pylori*) or cancer cases (for HPV), the RRs or ORs and attributable percentages used in our analyses.

2.2. Future infection prevalence

We assumed a constant prevalence of HBV (from 2007–2011) and *H. pylori* (from 1999–2000) to 2027. We projected the future chronic HCV prevalence based on prevalence at three time points (1999, 2004, and 2009). Chronic HCV prevalence at the three time points was estimated by weighting the available five-year birth cohort data,[90] by Canada's population to obtain the weighted average prevalence for Canadians aged 15 to 70. To project the future chronic HCV prevalence, an exponential regression was fit between the prevalence and the three time points.

For the baseline HPV prevalence projections, we also assumed no change in prevalence given the lack of evidence in support of an increasing or decreasing trend in the prevalence within cases. Although the prevalence of HPV within oropharyngeal cancer has increased over time,[89] mostly due to a decrease in cigarette smoking, we assumed that this trend would not continue post-2018.

2.3. Counterfactual scenarios

We projected the impact of four counterfactual scenarios: no change in the prevalence of HBV and *H. pylori* and a continuing trend for HCV, as well as 10%, 25% and 50% reductions in infection prevalence. These reductions were selected to respectively represent plausible minor, moderate and major prevalence reductions. The counterfactuals were "implemented" in the year 2018 with a 15-year latency to observe an impact on cancer incidence starting in 2033.

There is no treatment for HPV infection; it can be cleared by the immune system rather than by an intervention.[155] We purposely ignored the impact of cervical cancer screening in achieving further cervical cancer incidence reduction and thus selected counterfactuals based on HPV vaccination coverage in girls only, and in girls and boys. Canada's National Advisory

Committee on Immunization recommended HPV vaccination for girls in 2007 and for boys in 2012.[433] We considered several plausible counterfactuals for school-based HPV vaccination starting in 2018: 1) maintenance of current coverage, 2) decrease in coverage among girls (40%, 50%, and 60% coverage) and 3) increase in coverage among girls only to 80% and 4) an 80% coverage of school-aged girls and boys (which is sufficient for the elimination of HPV16 [434]), as both direct effects (e.g. those who were vaccinated are protected and no one else) and then as herd effects (e.g. vaccine protection extends beyond those directly immunized). Decreasing coverage was considered for two reasons. First, a 50% coverage, although lower than the national average, is the current level of coverage in certain regions of Canada.[8] In addition, some countries such as Denmark and Japan have experienced substantial decreases in the level of coverage due to unconfirmed reports of adverse events.[435,436] For example, in Sapporo, Japan, the reported three-dose HPV vaccination completion rates ranged from 68.4 to 74.0% and two years later it dropped to 0.6%.[436] For comparison, we also present the expected cancer incidence that could have occurred had the HPV vaccine has never been administered at any point in time.

2.4. Latency period

HBV, HCV, and *H. pylori* are associated with prolonged latencies that can span decades before cancer diagnosis.[82,96] For these infections, we assumed a 15-year interval between the time of prevalence reduction and its impact on the incidence of associated cancers; a shorter latency was an appropriate approach given that the data captured prevalent (recent and persistent) rather than incident infections. For HCV, the available data did not allow for the direct estimation of the prevalence among those 70 years of age or older, so we allowed for a longer latency (between 15 to 20 years) in this age range. For HPV-associated cancers, we did not account for a latency period because we utilized a cohort approach in which five-year age group cohorts (i.e., 20–24, 25–29, etc.) were followed through time to 2042.

2.5. Human papillomavirus model parameters

2.5.1 Start date of vaccine coverage

School-based immunization of girls in grades 4 to 7 was introduced in Canadian provinces from 2007 to 2010. Specifically, Ontario (Canada's most populous province) started vaccinating grade 7 girls in 2007, whereas Quebec began vaccinating grade 4 girls in 2008 and British

Columbia started vaccinating grade 6 girls in 2008.[8] We selected the year 2008 as the single start date for country-wide vaccination of girls, corresponding to the median year vaccination began. School-based catch-up HPV vaccination programs were extended to boys, first in Prince Edward Island (province with the smallest population) in 2013 and, to a few other jurisdictions (province/territory) in the following years. As we are not considering catch-up vaccination here, we did not consider the impact of catch-up vaccination targeted at boys prior to 2018.

2.5.2. Current vaccine coverage

To estimate current Canada-wide vaccine coverage across jurisdictions, we calculated a weighted proportion based on average vaccine completion rates (receiving the last dose of a two or three dose schedule) for the available school years within each jurisdiction.[8] The weights were represented by the proportion of girls aged 10–14 in a particular jurisdiction relative to their Canadian counterparts for the year 2014. The weights were based on the 2014 population levels because vaccine completion rates were reported for school years ranging from 2011/12 to 2015/16. Country-wide coverage was estimated because we lacked provincial level cancer incidence data for some HPV-associated cancer sites (e.g., vulva, vagina, base of tongue and tonsil), and provincial cancer incidence could only be projected to 2038 due to smaller sample size hindering stable projections past 2038. We calculated the school-based vaccination completion rate for Canada using a weighted mean based on the size of each province's proportion of the female Canadian population aged 10–14 years as weights. The resulting estimate, 72.4% among girls, was imputed to 2008, which was approximately the median year when school-based programs were introduced.

2.5.3. Vaccine efficacy and protection

Efficacy against high-grade cervical, vaginal, and vulvar disease/cancer based on per-protocol analyses of HPV vaccination trials was reported to range from 95 to 100% in HPV-naïve populations.[20,21] To be conservative, we used 95% efficacy in our calculations. Currently, three HPV vaccines are available;[437] the cancer causing HPV types covered by these vaccines are 16 and 18 (bi/quadrivalent and nonavalent), and the nonavalent also protects against types: 31, 33, 45, 52 and 58. Since the nonavalent vaccine will be in use in all Canadian jurisdictions as of 2018,

we modeled its use starting in 2018 for the other counterfactuals. For cohorts vaccinated prior to 2018, we assumed that the quadrivalent vaccine was administered.

With respect to cervical cancer, we utilized protection levels of 70.8% for the quad/bivalent and 89.5% for the nonavalent vaccines since these proportions represent the estimated relative contribution of the HPV types covered by the respective vaccines.[73] Since we had previously estimated the proportion of anogenital cancers due to high-risk HPV types (Table 1), we calculated the proportion of high-risk HPV types included in the vaccines to determine their associated level of protection. For this estimation we relied on data from a study that reported HPV type distribution in anogenital cancer specimens obtained from population-based registries in the United States.[438] Specifically, to determine the level of vaccine protection, we estimated the proportion of the identified high-risk types covered by the quad/bivalent and nonavalent vaccines. For example, because 91.2% of anal cancer specimens tested positive for a high-risk HPV type of which 89.1% were positive for HPV16/18 and 97.7% were positive for any of the HPV types covered by the nonavalent vaccine,[438] we accounted for protection levels in anal cancers of 89.1% for the quadrivalent and 97.7% for the nonavalent vaccines.

2.5.4. Herd immunity

The HPV vaccine confers different levels of herd immunity among non-vaccinated girls and boys. We extracted and interpolated herd effects from a modelling study that meta-analyzed transmission-dynamic models from high-income countries.[434] Brisson et al. calculated that 40% vaccine coverage of girls would produce 53% protection among women and 36% among men whereas for 80% coverage of girls, 93% protection among women and 83% among men would be observed.[434] For 50%, 60% and 72.4% vaccine coverage levels, we assumed that the herd effects would increase by 10% increments. For example, a 50% coverage of girls would produce an estimated effect of 63% (10% higher than the 53% herd effect reported for 40% coverage of girls) and a 46% coverage of boys (10% higher than the 36% herd effect for boys when 40% of girls are vaccinated). For current coverage of 72.4%, we increased the herd effect by an additional 2.4% to match the increase in coverage from 60% to 72.4%.

2.5.5. Estimating preventable cases

To determine the proportion of future cancer incidence that could be prevented under the different HPV vaccine coverage counterfactuals, we multiplied the following parameters: (1) proportion of cancer attributable to high-risk HPV types for anogenital cancers (ranging from 39.4% for penile cancer to 100.0% for cervical cancer – Table 1) and to HPV16 for head and neck cancers (ranging from 8.2% for oral cavity cancer to 60.2% for oropharyngeal cancer), (2) level of direct (40.0%–80.0%) or herd (36.0%–100.0%) vaccine coverage, (3) level of protection offered by the vaccines (70.8%–97.7%), and (4) vaccine efficacy (95.0%). The resulting proportion was then multiplied by the projected number of cancers to calculate the number of preventable cancers.

2.6. Cancer incidence

Supplementary Table 1 describes the modelling approach to estimate future cancer incidence (2018–2042) for each cancer. The projected number of cancers was estimated using three methods. The first method involved fitting different models with the 'Canproj' R package; this process is described in detail elsewhere.[439] The second involved applying a proportion to the Canproj projected cancer incidence to obtain the number of incident cancers for specific subsites. For example, this approach was utilized to determine the proportion of tongue cancer that is expected to be from the base of tongue and the proportion of stomach cancer that is expected to be from the non-cardia part of the stomach (Supplementary Table 1). Cancer incidence data for rare or subsite cancers were only available for two age groups (<50 and ≥50 years). To approximate the number of cancers occurring in five-year age groups, we partitioned the counts in these two age groups by the five-year age distributions from other related cancers. Specifically, the cervical cancer five-year age distribution within the <50 and ≥50 age groups was used to partition vaginal and vulvar cancers, and the tongue cancer five-year age distribution was used to partition tonsillar cancer, thereby allowing us to assess the impact of HPV vaccination on cancer incidence. As herd effects from girl-only vaccination do not confer protection among men who have sex with men (MSM), we estimated the proportion of anal cancers occurring among MSM. We calculated a proportion of 49.4% of anal cancers attributable to MSM by utilizing a RR

of 17.3 for the association between sexual orientation and anal cancer and a 6.0% prevalence of MSM among those aged 15 to 44 in the United States.[440,441]

2.7. Statistical analysis

The calculation of attributable risks has been previously published.[431] Briefly, to estimate the proportion of cancer that is attributable to HPV individual studies were pooled with a random effects model. A fixed effect model was used to produce a pooled measure of association between *H. pylori* and non-cardia gastric carcinoma. Meta-analyses were performed, and figures were produced in Stata v14 (StataCorp., College Station, TX, USA). R (version 3.4.1) was used to calculate the future preventable burden of HBV, HCV, and *H. pylori* associated cancers[112] and an electronic spreadsheet was used to estimate the future preventable burden of HPV associated cancers.

Ethics approval was granted for this project by the Health Research Ethics Board of Alberta - Cancer Committee (HREBA.CC-14-0220_REN4) and McGill University granted an exemption to research ethic board review.

3. RESULTS

A 50% reduction in HBV, HCV, *H. pylori* prevalence and 80% HPV vaccine coverage of girls and boys in 2018 resulted in an estimated 15,946 cancers that could be prevented from 2018 to 2042 (**Tables 2–5**). **Figures 1** and **2** demonstrate how the cumulative number of preventable cases increases over time, and for HBV, HCV and *H. pylori* after a latency period. A 50% reduction in the prevalence of HBV, HCV and *H. pylori* and 80% HPV coverage among girls and boys, could prevent an estimated 1.0% of all cancers diagnosed among men and 0.9% diagnosed among women in 2042 (data not shown).

3.1. Hepatitis B and C viruses

The future prevalence of HBV remained constant, however, the future prevalence of HCV was projected as steadily decreasing to 2042. A 50% reduction in the prevalence of HBV and HCV would result in slightly fewer projected hepatocellular carcinoma cases in 2042; from 3358 to 3210 for HBV and from 3358 to 3106 for HCV (Table 2). Cumulatively from 2018 to 2042, a 10% reduction in the prevalence of HBV and HCV would prevent 356 hepatocellular carcinomas as compared to a 50% prevalence reduction that would prevent 1782 hepatocellular carcinomas.

3.2. Helicobacter pylori

A 50% prevalence reduction in *H. pylori* would lead to fewer projected non-cardia gastric cancers (3579 cases) in 2042 compared to a no change in prevalence (5097 cases); and, fewer gastric mucosa associated lymphoid tissue (MALT) lymphomas, from 2403 to 1822 (Table 3). Cumulatively from 2018 to 2042, a 10% reduction in the prevalence of *H. pylori* would prevent 1749 non-cardia gastric cancers and gastric MALT lymphoma cases as compared to a 50% prevalence reduction that would result in 8744 fewer cases.

3.3. Human papillomavirus

If the estimated current Canada-wide HPV vaccine coverage of girls continued (72.4% direct coverage, but due to herd effects becoming equivalent to 85.4% coverage in girls and 68.4% in boys), an estimated total of 3976 anogenital cancers could be prevented from 2018 to 2042 (Table 4). The majority (85.4%) of these, preventable cases were cervical cancers and virtually all preventable cases occurred among women (99.4%). In contrast, the continuation of current HPV vaccine coverage could prevent more head and neck cancers among men (829 cases) than women (244 cases) from 2018 to 2042 (Table 5). Among all HPV-caused cancers, 80% vaccine coverage of girls and boys could prevent 4434 cancers among women and 928 among men by 2042 (Tables 4 and 5) among those less than age 45.

4. DISCUSSION

4.1. Hepatitis B and C viruses

The World Health Organization developed a global strategy to eliminate viral hepatitis with a focus on HBV and HCV by 2030;[442] Canada is a signatory to this strategy. For HBV, the major prevention measure is vaccination, which began as early as 1982 in Canada.[9] The Canadian government encourages health care providers to assess HBV status and immunize persons immigrating to Canada,[9] although this immunization does not appear to be systematic. The future incidence of hepatocellular carcinoma would be impacted by school- or infant-based universal immunization making a 50% reduction in the prevalence plausible. Approximately 12% of hepatocellular carcinoma cases could be prevented in 2042 with a 50% reduction in the prevalence of the hepatitis viruses in 2018; a 10% reduction would prevent only 2.4% of hepatocellular carcinoma cases in 2042. However, incorporating a 15-year latency for HBV and

HCV provided only a 10-year window (from 2032–2042) where cancer incidence could be changed by the prevalence reductions.

4.2. Helicobacter pylori

H. pylori was the infectious agent responsible for the most preventable cancer cases from 2018 to 2042 (8744 cancers with a 50% prevalence reduction). Although *H. pylori* is associated with a prolonged latency thereby expanding the opportunity to detect and deliver quadruple antibiotic therapy, there are challenges around determining who needs to be screened and concerns over increasing antibiotic resistance.[116] A 50% prevalence reduction may be more aspirational than attainable; however, the more achievable 25% prevalence reduction prevents more than 4,000 cancers from 2018 to 2042. When we projected the future prevalence of *H. pylori*, we assumed a constant trend, a decreasing trend in its prevalence would have resulted in fewer prevented cases whereas an increasing trend would have resulted in more.

4.3. Human papillomavirus

With 40% vaccination coverage of girls (herd effects lead to 53% coverage equivalents among girls and 36% among boys) achieved a notable number of preventable cases, with 3,503 potentially preventable cancers from 2018 to 2042. Since we used a birth cohort approach, the first two five-year cohorts were vaccinated prior to the application of counterfactual vaccine coverage in 2018, and thus the counterfactuals' impact on cancer incidence in these two cohorts was not modeled.

By projecting cancer incidence to only 2042, the first cohort of girls vaccinated in 2008 at ages 10 to 14 was then aged 40 to 44 in 2042, therefore only cancer incidence among individuals up to age 45 could be impacted. For boys, this constraint was even more pronounced as the vaccine was assumed to have been delivered starting in 2018. This restriction greatly influenced our results since only the first two cohorts could be followed to ages 35 to 44, whereas the remaining cohorts could only be followed to ages 30 to 34. Specifically, the impact of HPV vaccination counterfactuals was confined to cancers occurring in individuals under age 35 in 2042 and therefore differences between the counterfactual interventions are minimized as these only apply to younger cohorts. The impact of HPV vaccine coverage was limited to cancers occurring among individuals less than age 45, yet the majority of HPV-related cancers occurred in

individuals over age 45. Hence, our analysis provided a short-term assessment of the impact of school-based vaccination on cancer incidence among young Canadians.

Modeling the impact of HPV vaccine coverage counterfactuals involved several assumptions. First, the estimated herd effects relied on informed assumptions about the level of protection among non-vaccinated individuals as they were taken from a recent study for 40% and 80% direct coverage[434] but had to be interpolated for other coverage levels modeled here (i.e. 50%, 60%, 70%). Second, we used a more conservative approach to estimate current country-wide vaccine coverage by utilizing data on the completion of the recommended number of doses; yet, one dose has been shown to offer considerable protection against HPV-related diseases.[443] Third, we assumed that the vaccine confers long-term protection (up to 30 years in our calculations) against the HPV types it protects against.

There are several limitations of our analysis. First, we did not account for immigration in our calculations; for example, new arrivals who were not vaccinated through the school-based or catch-up vaccination programs are not covered by the counterfactuals and have a greater risk of developing HPV-associated cancers than the remaining Canadian population; however, herd effects are anticipated to minimize this concern. Second, although our estimate of country-wide vaccination was conservative (72.4%), there is substantial variation in the level of HPV vaccine coverage, hence some Canadian jurisdictions might not realize the reductions in cancer incidence that are possible with the counterfactual coverage levels. For example, receiving the recommended number of vaccine doses ranges from approximately 50% in Nunavut to 90% in Newfoundland and Labrador.[8] Conservatively, the impact of catch-up vaccination was not modeled, yet would result in more preventable cancers in the future. Third, improvements in cervical cancer screening technology coupled to vaccination coverage are likely to result in improved and more efficient cervical cancer prevention in the future, potentially leading to elimination of this disease.[437,444] Finally, we focused our analysis on the four infections that cause the most cancers in Canada and for which there are proven prevention strategies; however, other infections such as Epstein-Barr virus and human immunodeficiency virus also cause cancer and a reduction in their prevalence could lessen the future infection-associated cancer burden.

4.4. Implications for cancer prevention

With an aging population, the future burden of cancer in Canada is expected to substantially increase to 2032.[119] Changes in cancer risks due to major risk factors such as infections will have varying impacts on the future burden of cancer;[119] we identified the impact that four preventable and/or treatable infections can have on the future cancer burden. Even the short-term view presented here, reveals that different interventions have differing impacts on future incidence.

5. CONCLUSION

By modelling the impact of 10%, 25%, and 50% relative reductions in the prevalence of infections – HBV, HCV, and *H. pylori* – we estimated that more than 10,000 cancers could be prevented from 2018 to 2042 with a 50% prevalence reduction. The impact of 80% school-based HPV vaccine coverage among girls and boys would potentially prevent 5360 cancer cases from 2018 to 2042. Despite only capturing the impact of school-based HPV vaccination on cancers occurring among those less than age 45, our results indicate that increases in HPV coverage can result in meaningful decreases in HPV-related cancer incidence. With Canada's current cancer prevention resources, there is a substantial opportunity to reduce the future infection-associated cancer burden.

Table 1. Cancer types and proportions attributable to carcinogenic infections with modifiable prevalence in Canada[a]

Infection Cancer sites (ICD-O-3 codes)	**Prevalence of the infection in the population, %** [b]	**Odds ratio or relative risk**	**Attributable, %** [c]	
			Men	**Women**
Hepatitis B virus (HBV), *chronic infection*				
Hepatocellular carcinoma (C22, 817)	0.54 (men) 0.36 (women)	20.3	9.5	6.5
Hepatitis C virus (HCV), *chronic infection*				
Hepatocellular carcinoma (C22, 817)	1999: 1.09 (men) and 0.73 (women) 2004: 1.05 (men) and 0.70 (women) 2009: 0.99 (men) and 0.66 (women)	23.4	16.0	11.3
Non-Hodgkin lymphoma (9591)		1.35	0.3	0.2
Helicobacter pylori (H. pylori)				
Gastric non-cardia (C16.1–16.9)	18.0 (men) 17.2 (women)	9.4	60.0	59.0
Gastric MALT lymphoma (9699)		6.3	48.8	47.7
	Prevalence of the infection in cancer cases, %			
Human papillomavirus (HPV), high-risk types[d]				
Cervix (C53)	100.0	Not applicable as the prevalence in cancer cases approximates the proportion attributable to the infection.		
Anus (C21)	86.1 (men) 94.5 (women)			
Penis (C60)	39.4			
Vagina (C52)	72.2			
Vulva (C51)	76.8 (aged 18–49 years) 43.2 (aged ≥50 years)			
Human papillomavirus (HPV), type 16				
Oropharynx (C10, C01, C09) [e]	60.2			
Oral cavity (C03, C04, C02) [f]	8.2			
Larynx (C32)	12.7			

MALT = mucosa-associated lymphoid tissue

a. Detailed description of the prevalence and relative risk estimates can be found in Volesky et al., 2019 [(431)].

b. The prevalence of the infection in the population was calculated by weighting the age-group specific prevalence estimates by the Canadian population for each sex.

c. The attributable percent by sex was calculated by dividing the number of attributable cases by the number of cases, and hence it does not reflect the proportion attributable by specific age groups.

d. High-risk HPV types include types classified by the International Agency for Research on Cancer as Group 1 (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59), Group 2A (68) and Group 2B (34, 53, 66, 70 and 73) carcinogens. HPV97 was also considered high-risk types.

e. Oropharynx subsites: oropharynx (C10), base of the tongue (C01), and tonsil (C09).

f. Oral cavity subsites: gum (C03), floor of mouth (C04), other and unspecified parts of tongue (C02).

Table 2. Projected number of cancer cases and proportions attributable to chronic hepatitis B and C viruses by sex that could be prevented in 2042 under different counterfactuals

Sex	Future burden measures	No change	Cancer burden by reductions in infection prevalence		
			10%	25%	50%
Hepatitis B virus, Hepatocellular carcinoma					
Men	Projected in 2042	2640	2615	2578	2516
	PIF, (%)	--	0.9	2.4	4.7
	Prevented in 2042	0	25	62	125
	Prevented 2018–2042	0	110	275	551
Women	Projected in 2042	718	713	706	695
	PIF, (%)	--	0.6	1.6	3.2
	Prevented in 2042	0	5	12	23
	Prevented 2018–2042	0	20	50	100
Both	Projected in 2042	3358	3329	3284	3210
	PIF, (%)	--	0.9	2.2	4.4
	Prevented in 2042	0	30	74	148
	Prevented 2018–2042	0	130	326	651
Hepatitis C virus, Hepatocellular carcinoma					
Men	Projected in 2042	2640	2598	2535	2429
	PIF, (%)	--	1.6	4.0	8.0
	Prevented in 2042	0	42	106	212
	Prevented 2018–2042	0	190	476	952
Women	Projected in 2042	718	710	697	677
	PIF, (%)	--	1.1	2.8	5.6
	Prevented in 2042	0	8	20	41
	Prevented 2018–2042	0	36	90	179
Both	Projected in 2042	3358	3308	3232	3106
	PIF, (%)	--	1.5	3.8	7.5
	Prevented in 2042	0	50	126	252
	Prevented 2018–2042	0	226	565	1131
Hepatitis C virus, Non-Hodgkin lymphoma					
Men	Projected in 2042	5850	5849	5846	5842
	PIF, (%)	--	0.0	0.1	0.1
	Prevented in 2042	0	2	4	9
	Prevented 2018–2042	0	8	19	39
Women	Projected in 2042	4750	4749	4748	4745
	PIF, (%)	--	< 0.1	< 0.1	0.1
	Prevented in 2042	0	1	2	5
	Prevented 2018–2042	0	4	10	21
Both	Projected in 2042	10,600	10,598	10,594	10,587
	PIF, (%)	--	0	0.1	0.1
	Prevented in 2042	0	3.0	7.0	13.0
	Prevented 2018–2042	0	12	30	59
Hepatitis C virus, Total					
Men	Projected in 2042	8491	8447	8381	8271
	PIF, (%)	--	0.5	1.3	2.6
	Prevented in 2042	0	44	110	220
	Prevented 2018–2042	0	198	495	990
Women	Projected in 2042	5468	5459	5445	5423
	PIF, (%)	--	0.2	0.4	0.8
	Prevented in 2042	0	9	23	45
	Prevented 2018–2042	0	40	100	200
Both	Projected in 2042	13,959	13,905	13,826	13,693
	PIF, (%)	--	0.4	1.0	1.9
	Prevented in 2042	0	53	133	265
	Prevented 2018–2042	0	238	595	1190

PIF = potential impact fraction

Table 3. Projected number of cancer cases and proportions attributable to *Helicobacter pylori* that could be prevented in 2042 under different counterfactuals

Sex	Future burden measures	No change	Cancer burden by reductions in infection prevalence 10%	25%	50%
Gastric MALT lymphoma					
Men	Projected in 2042	1389	1321	1219	1050
	PIF, (%)	--	4.9	12.2	24.4
	Prevented in 2042	0	68	170	339
	Prevented 2018–2042	0	272	679	1,358
Women	Projected in 2042	1014	966	893	772
	PIF, (%)	--	4.8	11.9	23.8
	Prevented in 2042	0	48	121	242
	Prevented 2018–2042	0	198	496	992
Both	Projected in 2042	2403	2287	2112	1822
	PIF, (%)	--	4.8	12.1	24.2
	Prevented in 2042	0	116	290	581
	Prevented 2018–2042	0	470	1,175	2,351
Gastric non-cardia cancer					
Men	Projected in 2042	2823	2654	2399	1976
	PIF, (%)	--	6.0	15.0	30.0
	Prevented in 2042	0	170	424	848
	Prevented 2018–2042	0	717	1,792	3,585
Women	Projected in 2042	2274	2140	1939	1604
	PIF, (%)	--	5.9	14.7	29.5
	Prevented in 2042	0	134	335	670
	Prevented 2018–2042	0	562	1404	2809
Both	Projected in 2042	5097	4794	4338	3579
	PIF, (%)	--	6.0	14.9	29.8
	Prevented in 2042	0	304	759	1518
	Prevented 2018–2042	0	1279	3197	6393
Total					
Men	Projected in 2042	4212	3975	3619	3025
	PIF, (%)	--	5.6	14.1	28.2
	Prevented in 2042	0	237	593	1187
	Prevented 2018–2042	0	989	2472	4943
Women	Projected in 2042	3288	3105	2832	2376
	PIF, (%)	--	5.5	13.9	27.7
	Prevented in 2042	0	182	456	912
	Prevented 2018–2042	0	760	1,900	3,801
Total	Projected in 2042	7500	7080	6450	5401
	PIF, (%)	--	5.6	14.0	28.0
	Prevented in 2042	0	420	1049	2099
	Prevented 2018–2042	0	1749	4372	8744

PIF = potential impact fraction, MALT = mucosa-associated lymphoid tissue

Table 4. Projected number of cancer cases and proportions attributable to human papillomavirus and the proportion of anogenital cancer cases that could be prevented in 2042 according to variations in school-based HPV vaccine coverage in Canada[a,b,c]

Cancer site, sex	Future burden measures	Effect:	Lower (%)							Current (%)		Higher (%)		
			Direct[d]	Direct	Herd	Direct	Herd	Direct	Herd	Direct	Herd	Direct	Herd	Herd
		Girls:	0.0	40.0	53.0	50.0	63.0	60.0	73.0	72.4	85.4	80.0	93.0	100.0
		Boys:	0.0	0.0	36.0	0.0	46.0	0.0	56.0	0.0	68.4	0.0	83.0	100.0
Cervix	Projected in 2042		1939	1723	1684	1693	1654	1663	1624	1626	1587	1603	1564	1543
	Prevented in 2042		0	216	255	246	285	276	315	313	352	336	375	396
	Prevented 2018–2042		0	2813	2980	2941	3108	3070	3236	3228	3395	3326	3492	3583
Anus, men[e]	Projected in 2042		345	345	345	345	345	345	344	345	344	345	344	343
	Prevented in 2042		0	0	0	0	0	0	0	0	1	0	1	1
	Prevented 2018–2042		0	0	8	0	9	0	10	0	11	0	13	25
Anus, women	Projected in 2042		775	758	757	757	756	756	755	755	754	755	754	753
	Prevented in 2042		0	17	18	18	19	19	20	20	21	20	21	22
	Prevented 2018–2042		0	130	131	131	132	132	133	133	135	134	136	136
Anus, both	Projected in 2042		1120	1103	1102	1102	1101	1101	1100	1100	1099	1100	1098	1097
	Prevented in 2042		0	17	18	18	19	19	20	20	21	20	22	23
	Prevented 2018–2042		0	130	139	131	141	132	143	133	146	134	148	161
Penis	Projected in 2042		260	260	258	260	258	260	258	260	258	260	258	258
	Prevented in 2042		0	0	0	2	0	2	0	2	0	2	0	2
	Prevented 2018–2042		0	0	13	0	14	0	14	0	14	0	15	15
Vagina	Projected in 2042		172	168	167	167	167	167	166	166	165	166	165	164
	Prevented in 2042		0	4	5	5	5	5	6	6	7	6	7	8
	Prevented 2018–2042		0	52	55	55	58	57	61	60	64	62	66	69
Vulva	Projected in 2042		987	964	960	961	956	957	953	953	949	951	947	945
	Prevented in 2042		0	23	27	27	31	30	34	34	38	36	40	42
	Prevented 2018–2042		0	296	313	309	327	323	341	340	358	350	368	384
Total, women	Projected in 2042		3873	3613	3568	3578	3533	3544	3499	3501	3456	3475	3430	3406
	Prevented in 2042		0	260	305	295	340	330	374	372	417	399	443	468
	Prevented 2018–2042		0	3291	3480	3436	3625	3582	3771	3762	3951	3873	4062	4174
Total, men	Projected in 2042		605	605	603	605	603	605	603	605	603	605	602	602
	Prevented in 2042		0	0	2	0	2	0	2	0	2	0	2	3
	Prevented 2018–2042		0	0	21	0	22	0	24	0	25	0	27	40
Total, both	Projected in 2042		4478	4218	4171	4183	4136	4149	4102	4106	4059	4080	4032	4007
	Prevented in 2042		0	260	307	295	342	330	376	372	419	399	446	471
	Prevented 2018–2042		0	3291	3501	3436	3648	3582	3794	3762	3976	3873	4089	4213

a. We did not round numbers when performing the analysis and hence some figures do not add up.
b. The direct effects of 80% vaccine coverage among boys was not modeled.
c. Since cancer incidence was projected to only 2042, the first vaccinated cohort of girls vaccinated in 2008 at ages 10–14 were aged 40–44 in 2042 meaning that only cancer incidence among those up to age 45 could be impacted by vaccination.
d. Direct effects of 0.0 among girls and boys assume that the HPV vaccination was never administered at any point in time in Canada.
e. We estimated that 49.4% of anal cancers occur among men who have sex with men and hence are not impacted by herd effects of girls only vaccination.

Table 5. Projected number of cancer cases and proportions attributable to human papillomavirus and the proportion of head and neck cancer cases that could be prevented in 2042 with various changes in school-based HPV vaccine coverage in Canada[a,b,c]

Cancer site, Sex	Future burden measures	Effect:	Lower (%)							Current (%)		Higher (%)		
			Direct[d]	Direct	Herd	Direct	Herd	Direct	Herd	Direct	Herd	Direct	Herd	Herd
		Girls:	0.0	40.0	53.0	50.0	63.0	60.0	73.0	72.4	85.4	80.0	93.0	100.0
		Boys:	0.0	0.0	36.0	0.0	46.0	0.0	56.0	0.0	68.4	0.0	83.0	100.0
Oropharynx, men [e]	Projected in 2042		3469	3469	3363	3469	3360	3469	3356	3469	3352	3469	3348	3342
	Prevented in 2042		0	0	106	0	109	0	113	0	117	0	121	127
	Prevented 2018–2042		0	0	742	0	760	0	778	0	800	0	826	857
Oropharynx, women [e]	Projected in 2042		914	894	892	892	890	891	889	889	887	888	885	884
	Prevented in 2042		0	20	23	22	24	24	26	26	28	27	29	30
	Prevented 2018–2042		0	186	193	192	199	198	205	205	213	210	217	227
Oropharynx, both [e]	Projected in 2042		4383	4363	4255	4361	4250	4360	4245	4358	4239	4357	4233	4226
	Prevented in 2042		0	20	129	22	134	24	138	26	145	27	151	157
	Prevented 2018–2042		0	186	935	192	959	198	983	205	1013	210	1044	1084
Oral cavity, men	Projected in 2042		761	761	760	761	760	761	759	761	759	761	759	759
	Prevented in 2042		0	0	2	0	2	0	2	0	2	0	2	2
	Prevented 2018–2042		0	0	16	0	16	0	17	0	17	0	18	19
Oral cavity, women	Projected in 2042		858	856	856	856	856	856	856	856	855	855	855	855
	Prevented in 2042		0	2	2	2	2	2	3	3	3	3	3	3
	Prevented 2018–2042		0	22	24	23	24	24	25	25	26	26	27	28
Oral cavity, both	Projected in 2042		1619	1617	1616	1617	1615	1617	1615	1617	1615	1617	1615	1614
	Prevented in 2042		0	2	4	2	4	2	4	3	5	3	5	5
	Prevented 2018–2042		0	22	40	23	41	24	42	25	44	26	45	46
Larynx, men	Projected in 2042		1230	1229	1229	1230	1228	1230	1228	1230	1228	1230	1228	1228
	Prevented in 2042		0	0	2	0	2	0	2	0	2	0	2	2
	Prevented 2018–2042		0	0	11	0	12	0	12	0	12	0	12	12
Larynx, women	Projected in 2042		187	186	186	186	186	186	186	186	186	186	186	186
	Prevented in 2042		0	0	0	0	0	0	0	0	0	0	0	1
	Prevented 2018–2042		0	4	4	4	4	4	5	5	5	5	5	5
Larynx, both	Projected in 2042		1417	1415	1415	1417	1415	1416	1415	1416	1415	1416	1415	1415
	Prevented in 2042		0	0	2	0	2	0	2	0	2	0	2	2
	Prevented 2018–2042		0	4	16	4	16	4	16	5	16	5	17	17
Total, men	Projected in 2042		5460	5459	5351	5460	5348	5460	5344	5460	5340	5460	5335	5330
	Prevented in 2042		0	0	109	0	113	0	116	0	120	0	125	131
	Prevented 2018–2042		0	0	769	0	788	0	806	0	829	0	856	888
Total, women	Projected in 2042		1959	1937	1934	1935	1932	1933	1931	1931	1928	1929	1927	1926
	Prevented in 2042		0	23	25	25	27	26	29	29	31	30	33	34
	Prevented 2018–2042		0	212	221	219	228	226	235	235	244	240	249	260
Total, both	Projected in 2042		7420	7395	7285	7395	7280	7393	7275	7391	7268	7390	7262	7255
	Prevented in 2042		0	23	135	25	140	26	145	29	151	30	158	165
	Prevented 2018–2042		0	212	990	219	1016	226	1041	235	1073	240	1105	1148

a. We did not round numbers when performing the analysis and hence some figures do not add up.
b. Direct effects of 80% vaccine coverage among boys was not modeled.
c. Since cancer incidence was projected to only 2042, the first vaccinated cohort of girls vaccinated in 2008 at ages 10–14 were aged 40–44 in 2042 meaning that only cancer incidence among those up to age 45 could be impacted by vaccination.
d. Direct effects of 0.0 among girls and boys assume that the HPV vaccination was never administered at any point in time in Canada.
e. Included the base of the tongue and tonsils.

Fig. 1. Projected cumulative preventable cases attributable to hepatitis B and C viruses (A) and Helicobacter pylori (B) by applying counterfactual prevalence reductions

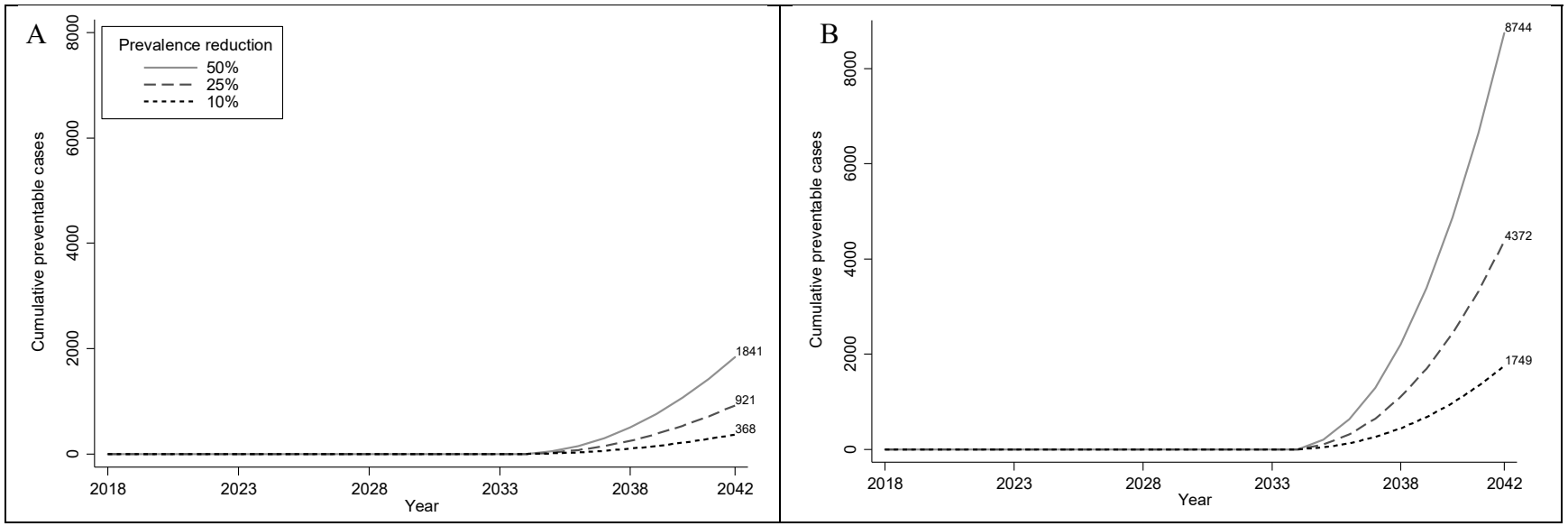
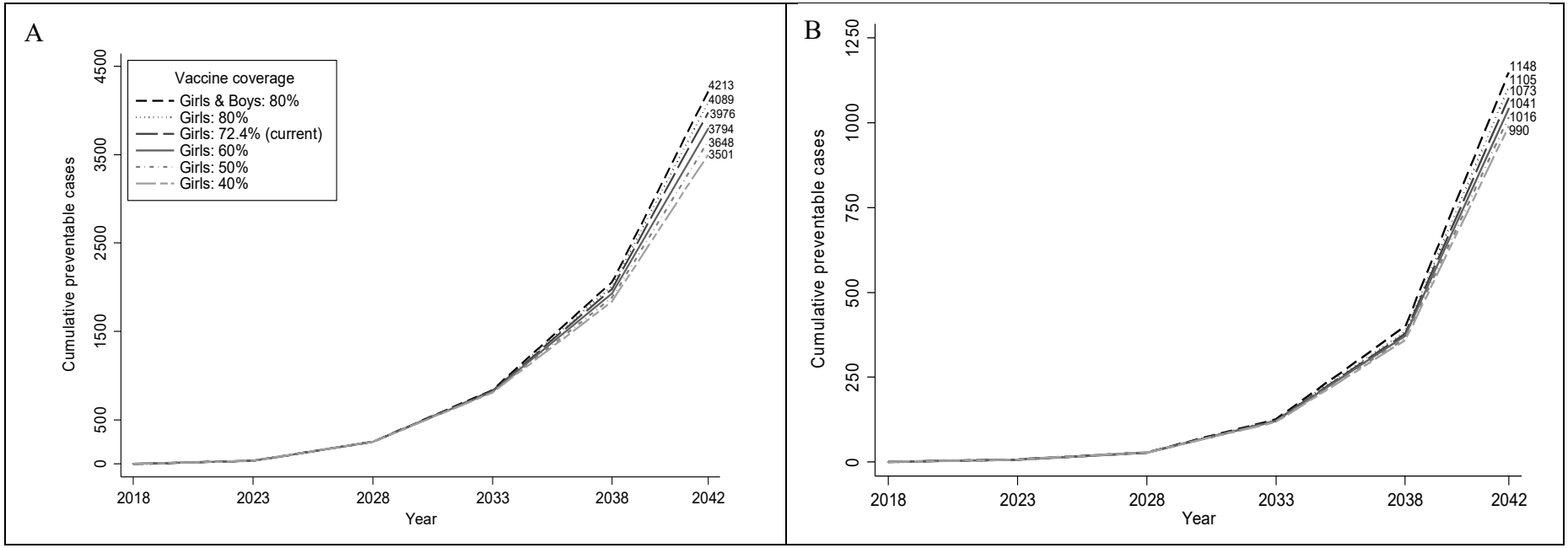


Fig. 2. Projected cumulative preventable anogenital cancers (A) and head and neck cancers (B) attributable to human papillomavirus by applying school-based HPV vaccine coverage counterfactuals[a,b]



a. The vaccine coverage level refers to the percent of those aged 10–14 receiving the HPV vaccine.

b. We modeled the herd effects of vaccine coverage (e.g. 40% coverage of girls produces 53% coverage of girls and 36% coverage of boys).

Supplementary Table 1. Model selection and description of how future cancer incidence was projected

Infection Cancer site(s)	Canproj model Description of any addition estimations required
Hepatitis B virus	
Hepatocellular carcinoma	Negative-binomial based age-drift-period-cohort Fixed proportion (0.797) was applied to liver cancer to obtain the number that would be hepatocellular carcinoma.[109]
Hepatitis C virus	
Hepatocellular carcinoma	Negative-binomial based age-drift-period-cohort Fixed proportion (0.797) was applied to liver cancer incidence to obtain the number that would be hepatocellular carcinoma.[109]
Non-Hodgkin lymphoma	Negative-binomial based age-drift-period-cohort
Helicobacter pylori	
Gastric MALT lymphoma	Poisson regression was fit between the incidence rate and the year (1999–2010). As the past incidence of this cancer was very low prior to 1999 we assumed that underreporting may have unduly influenced the reported incidence and used data starting in 1999 to avoid projecting an erroneous trend.
Gastric non-cardia cancer	Negative-binomial based age-drift-period-cohort (gastric cancer overall) Proportion of gastric cancer that is expected to be non-cardia gastric cancer was applied to the projected gastric cancer incidence. This proportion was calculated by averaging the proportion of gastric cancer that is classified as non-cardia over the last five years of available cancer incidence data (2011–2015 for Canada and 2006–2010 for the province of Quebec) and multiplying it by the projected overall gastric cancer incidence.
Human papillomavirus, anogenital	
Anus	Negative-binomial based age-drift-period-cohort (males), binomial based age-specific trend (females)
Cervix	Negative-binomial based age-drift-period-cohort
Penis	Negative-binomial based age-cohort trend
Vagina	Poisson regression fit between the incidence rate and the year (1992–2010)
Vulva	Poisson regression fit between incidence rate and the year (1992–2010)
Human papillomavirus, head and neck	
Oropharynx	Included subsites: base of the tongue (C01), oropharynx (C10), and tonsil (C09). Base of tongue: Negative binomial based age specific trend (tongue overall). A fixed proportion of tongue cancer that is expected to be base of tongue (BOT) was estimated for each year by calculating the log odds of tongue cancer being BOT from Canadian cancer incidence from 1992–2010, using logistic regression, separately by sex Oropharynx: Negative-binomial based age-specific trend Tonsil: Poisson regression between incidence rate and the year (1992–2010). We were only able to obtain tonsil cancer incidence by aggregated age groups (age <50, ≥50). To allow us to model the impact of HPV vaccination by birth cohorts, we estimated tonsillar cancer incidence by 5-year age groups with the 5-year age distribution of tongue cancer among those under age 50 and those over age 50 for each year from 2018–2042.
Oral cavity	Included subsites were: floor of mouth (C04), tongue (C02), gum and other mouth (C03, C06). Floor of mouth: Negative binomial based age specific trend Tongue (mobile): Negative binomial based age specific trend (tongue overall). The tongue cancer incidence remaining after subtracting the estimated base of tongue incidence (see above for how base of tongue incidence was estimated) approximated the mobile part of the tongue. Gum and other mouth: Negative-binomial based age specific trend (for males) and Poisson-based age-drift-period-cohort (for females)
Larynx	Negative-binomial based age-drift-period-cohort

Chapter 5: Discussion

The overarching goal of this thesis was to assess the impact of infections on North American cancer incidence. To that end, we estimated that 3.7% of the 189,530 cancers diagnosed among individuals aged ≥18 years in Canada in 2015 and 4.3% of the 1,662,102 cancers diagnosed among individuals aged ≥20 years in the US in 2017, were attributable to seven infections – HBV, HCV, *H. pylori*, EBV, HPV, HHV-8, and HTLV-1. While at first, these proportions may not seem substantial, they translated to 7097 cases in Canada and 71,469 in the US in a single year. Over several years the number of cancers due to infections would be far greater.

The majority of infection-attributable cancers (90.0% [n = 6389] in Canada and 86.5% [n = 61,799] in the US) were due to infections (HBV, HCV, *H. pylori* and HPV) where effective prevention and/or treatment interventions exist. This highlights the tremendous opportunity to accelerate uptake of available interventions. Of note, the burden of cancer due to infections represents one group of outcomes. The infections included in this analysis are also associated with several other negative health outcomes, such as peptic ulcers and gastritis (*H. pylori*), mononucleosis (EBV), and chronic liver disease and failure (HBV and HCV). Thus, interventions that can prevent or treat these infections can reduce the burden of cancer and other adverse health outcomes.

The estimates of the overall impact of infections on cancer incidence reported here (3.7% of cancers in Canada and 4.3% in the US) do not *substantially* differ from the estimates (3.6% for Canada and 4.8% for the US) reported by de Martel and colleagues in the most recent global analysis assessing the burden of infection-attributable cancers in 2018.[57] This is in spite of the current analysis utilizing a different approach to calculate PAFs for HBV, HCV, and *H. pylori*;[3] for the US specifically, including HBV and HCV and intra and extrahepatic bile duct cancer, *H. pylori* and esophageal adenocarcinoma, and EBV-related DLBCL and gastric cancer, and childhood

[3] Since the global analysis utilized the Miettinen formula (which combines exposure prevalence among cases and the relative risk) instead of the Levin formula utilized in this thesis research (Levin's formula combines exposure prevalence in the population and the relative risk), the types of data we used to source these estimates differed. The global analysis included studies of non-cardia gastric cancer that tested for *H. pylori* via immunoblot – we included studies utilizing immunoblot as well as those using ELISA/EIA (but corrected for measurement error).

cancers. Since country-level analyses have been published for few nations, it is reassuring that the estimates are comparable.

This thesis research was focused on Western populations,[45,445] yet the burden of infection-attributable cancers is much higher in low to middle income countries.[54-56,446] The vast difference in the proportion of cancers due to infections in high versus low income countries is driven by factors such as higher *H. pylori* prevalence related to crowding and sanitation conditions, relatively higher HIV prevalence, the presence of carcinogenic parasitic infections, limited access to HBV and HPV vaccination, a lack of access to cervical cancer screening, etc.[46,54,447] Within Canada and the US, the burden of infection-attributable cancers may be highly disparate between populations. While we generated estimates by sex and for the US by sex and age groups, this masks differences in the infection-attributable burden across different groups; in particular, among the immunocompromised. For example, it was estimated that 40% of cancers diagnosed in 2008 were attributable to infection among PLWH in the US.[227] This is in sharp contrast to the estimated 5.0% of cancers in the general US population the study authors estimated to be due to infections after adjusting the age and sex distribution to match that of HIV-positive population.

In the Canadian analysis, no specific adjustment was made to the cancer incidence data to account for potentially higher PAFs among PLWH. While data originating from the US and European countries existed to calculate PAFs for PLWH (for HPV in anal cancer and EBV in select lymphomas), we lacked estimates of the proportion of anal and lymphoma cancer incidence occurring among PLWH in Canada. Such estimates, would have allowed us to partition cancer incidence and apply separate PAFs for PLWH and the general population, like we did for the US.[236,237] In the US analysis, it was not necessary to adjust the number of anal cancers or Burkitt lymphomas attributable to infection because the individual studies included a mixture of HIV negative, PLWH, or for most studies HIV status unknown populations. Therefore, PLWH are accounted for to some extent in the resulting PAF. However, the adjustment was important for Hodgkin lymphoma and DLBCLs in the US because the PAFs greatly varied by HIV status. The Canadian analysis did not include DLBCLs but for Hodgkin lymphoma two of the four studies meta-analyzed for this cancer excluded PLWH and thus an underestimation in the proportion of

Hodgkin lymphomas due to EBV is possible. Note that, calculating separate PAFs by HIV status is reliant on individual studies reporting the HIV status of cases, but many studies did not report this information. In summary, although we did not account for HIV status in the Canadian analysis, the proportion of missed attributable cases is expected to be minimal.

While we included cancers beyond IARC's list of cancers classified as having 'sufficient' evidence (these cancers were laryngeal cancer [HPV], and for the US only: intra and extrahepatic bile duct cancer [HBV, HCV], DLBCL and gastric carcinoma [EBV], and esophageal adenocarcinoma [*H. pylori*]), there are other associations that may warrant inclusion in future analyses. Specifically, Merkel cell polyomavirus associated Merkel cell carcinoma, and HBV related NHL. In the North American context, these two additions would not be expected to substantially increase the infection-attributable cancer burden because, Merkel cell carcinoma is a rarer cancer (<3000 cases diagnosed in the US in 2017),[235] chronic HBV prevalence is low (<0.5% HBsAg positive in the weighted and imputed NHANES data), and the measure of association between HBV and NHL in Western populations is modest (pooled OR ≈ 1.6).[448] Nevertheless, such associations have the potential to increase the burden of cancers attributable to specific infections.

While the limitations of this work were described in the manuscripts, several are highlighted here. First, is the issue of sparse data. Data were particularly limited for *H. pylori* and gastric MALT lymphoma, EBV and Burkitt lymphoma, and HPV and vaginal cancer. Second, the Canadian analysis relied heavily on data collected in the US and Europe, and several infection-cancer pairs in the US analysis relied on data collected in Europe. This presumes that there are not differences in the prevalence of the infection in cancer cases in different regions. We assumed that measures of association (for HBV, HCV, and *H. pylori*) calculated from cohorts, nested case-controls, and case-control studies were transportable between Western countries. PAFs, even those presented here for Canada and the US, are not directly comparable because of methods differ ultimately hindering comparisons. Our modeling on the future burden of cancers attributable to HBV, HCV, and *H. pylori* were based on simple counterfactual scenarios and are therefore not tied to specific existing interventions.

We encourage future country-level analyses that seek to estimate the burden of infection-attributable cancers to (i) report their methodology for finding, selecting, and including

studies, (ii) where possible match infection prevalence in cancer cases estimates to cancer incident data (i.e., EBV prevalence in NPC for the US was calculated and thus we applied to NPC cancer incidence rather than nasopharynx overall), (iii) account for unspecified cancers (NOS, not classified, unspecified lymphomas) where the data permit, (iv) attempt to provide estimates for children/adolescents who are susceptible to EBV-related cancers but typically excluded, and (v) consider cancers that had 'limited' evidence for the role of the infection more than 10 years ago when the IARC working group met but for which data has accumulated since then.

Estimates of the impact infections have on cancer incidence require regular updating. There are many reasons updates are important, such as notable changes in the prevalence of infections (HCV post-direct acting antiviral therapy introduction), changes in the gold standard method for infection detection (as was seen for *H. pylori* measurement and NCGC), changing cancer incidence (which can include changes in the distribution of cancer subtypes more closely related to particular infections), changing population demographics, and the emergence of new carcinogenic infections (discovery of Merkel cell polyomavirus in 2008) and the accumulation of evidence for additional cancers related to group 1 infections.

Main conclusions

- Seven infections related to 20 different cancers, were responsible for 3.7% (7097) cancers diagnosed among those aged ≥18 years in Canada in 2015.
- Seven infections related to 26 different cancers, were responsible for 4.3% (71,469) cancers among individuals aged ≥20 years in the US in 2017.
- HPV was the most important infectious cause of cancer among adults in North America accounting for more than half the infection-attributable cancer burden.
- One infection, EBV, was responsible for 2.6% of cancers among those aged 0–19 in Europe and North America in 2020.
- There is potential to lessen the future burden of cancers related to infections.

Together the findings reported in the manuscripts and their accompanying supplements, provide a comprehensive portrait for the burden of infection-attributable cancers in North America.

References

1. International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Biological Agents. Lyon, France; 2012.

2. Bouvard V, Baan R, Straif K, Grosse Y*, et al.* A review of human carcinogens--Part B: biological agents. Lancet Oncol. 2009;10(4):321-2.

3. Javier RT, Butel JS. The history of tumor virology. Cancer Res. 2008;68(19):7693-706.

4. Franco EL, Rohan TE, Villa LL. Epidemiologic evidence and human papillomavirus infection as a necessary cause of cervical cancer. J Natl Cancer Inst. 1999;91(6):506-11.

5. Walboomers JM, Jacobs MV, Manos MM, Bosch FX*, et al.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 1999;189(1):12-9.

6. Marshall BJ. Unidentified curved bacili on gastric epithelium in active chronic gastritis (Letter to the Editor). Lancet. 1983;i:1273-5.

7. Hunt R, Fallone C, Veldhuyzan van Zanten S, Sherman P*, et al.* Canadian Helicobacter Study Group Consensus Conference: Update on the management of *Helicobacter pylori*--an evidence-based evaluation of six topics relevant to clinical outcomes in patients evaluated for *H pylori* infection. Can J Gastroenterol. 2004;18(9):547-54.

8. Shapiro GK, Guichon J, Kelaher M. Canadian school-based HPV vaccine programs and policy considerations. Vaccine. 2017;35(42):5700-7.

9. Government of Canada. Canadian Immunization Guide: Part 4 - Active Vaccines 2017 [Available from: https://www.canada.ca/en/public-health/services/publications/healthy-living/canadian-immunization-guide-part-4-active-vaccines/page-7-hepatitis-b-vaccine.html].

10. Geddawy A, Ibrahim YF, Elbahie NM, Ibrahim MA. Direct Acting Anti-hepatitis C Virus Drugs: Clinical Pharmacology and Future Direction. J Transl Int Med. 2017;5(1):8-17.

11. Xu J, Murphy SL, Kochanek KD, Arias E. Deaths: Final Data for 2019. National Vital Statistics Reports, Centers for Disease Control and Prevention. 2021;70(8).

12. Statistics Canada. Deaths, 2019: Government of Canada; 2020 [Available from: https://www150.statcan.gc.ca/n1/daily-quotidien/201126/dq201126b-eng.htm].

13. Canadian Cancer Statistics Advisory Committee in collaboration with the Canadian Cancer Society; Statistics Canada and the Public Health Agency of Canada. Canadian Cancer Statistics 2021. Toronto, ON; 2021.

14. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin. 2021;71(1):7-33.

15. Whiteman DC, Webb PM, Green AC, Neale RE*, et al.* Cancers in Australia in 2010 attributable to modifiable factors: summary and conclusions. Aust N Z J Public Health. 2015;39(5):477-84.

16. Brown KF, Rumgay H, Dunlop C, Ryan M*, et al.* The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. Br J Cancer. 2018;118(8):1130-41.

17. Islami F, Goding Sauer A, Miller KD, Siegel RL*, et al.* Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. CA Cancer J Clin. 2018;68(1):31-54.

18. Poirier AE, Ruan Y, Volesky KD, King WD*, et al.* The current and future burden of cancer attributable to modifiable risk factors in Canada: Summary of results. Prev Med. 2019;122:140-7.

19. Soerjomataram I, Shield K, Marant-Micallef C, Vignat J*, et al.* Cancers related to lifestyle and environmental factors in France in 2015. Eur J Cancer. 2018;105:103-13.

20. Future II Study Group. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. N Engl J Med. 2007;356(19):1915-27.

21. Huh WK, Joura EA, Giuliano AR, Iversen OE*, et al.* Final efficacy, immunogenicity, and safety analyses of a nine-valent human papillomavirus vaccine in women aged 16-26 years: a randomised, double-blind trial. Lancet. 2017;390(10108):2143-59.

22. Wedemeyer H, Craxi A, Zuckerman E, Dieterich D*, et al.* Real-world effectiveness of ombitasvir/paritaprevir/ritonavir+/-dasabuvir+/-ribavirin in patients with hepatitis C virus genotype 1 or 4 infection: A meta-analysis. J Viral Hepat. 2017;24(11):936-43.

23. Patel SV, Jayaweera DT, Althoff KN, Eron JJ*, et al.* Real-world efficacy of direct acting antiviral therapies in patients with HIV/HCV. PLoS One. 2020;15(2):e0228847.

24. International Agency for Research on Agency. IARC Monographs: World Health Organization; 2016 [Available from: http://monographs.iarc.fr/].

25. International Agency for Research on Agency. Malaria and Some Polyomaviruses (SV40, BK, JC, and Merkel Cell Viruses). Lyon, France; 2014.

26. International Agency for Research on Cancer. Human Papillomavirus. Lyon, France; 2007.

27. International Agency for Research on Cancer. Agents Classified by the IARC Monographs, Volumes 1–130: World Health Organization; 2021 [Available from: https://monographs.iarc.who.int/agents-classified-by-the-iarc/].

28. International Agency for Research on Cancer. IARC Monographs on the Identification of Carcinogenic Hazards to Humans: List of Classifications, 2021 [Available from: https://monographs.iarc.who.int/list-of-classifications/].

29. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science. 2008;319(5866):1096-100.

30. Bouvard V, Baan RA, Grosse Y, Lauby-Secretan B*, et al.* Carcinogenicity of malaria and of some polyomaviruses. Lancet Oncol. 2012;13(4):339-40.

31. Moore PS, Chang Y. Why do viruses cause cancer? Highlights of the first century of human tumour virology. Nat Rev Cancer. 2010;10(12):878-89.

32. Shannon-Lowe C, Rickinson A. The Global Landscape of EBV-Associated Tumors. Front Oncol. 2019;9:713.

33. de-The G, Day NE, Geser A, Lavoue MF*, et al.* Sero-epidemiology of the Epstein-Barr virus: preliminary analysis of an international study - a review. IARC Sci Publ. 1975(11 Pt 2):3-16.

34. Young LS, Rickinson AB. Epstein-Barr virus: 40 years on. Nat Rev Cancer. 2004;4(10):757-68.

35. Hooi JKY, Lai WY, Ng WK, Suen MMY*, et al.* Global Prevalence of *Helicobacter pylori* Infection: Systematic Review and Meta-analysis. Gastroenterology. 2017.

36. Kuipers EJ. Review article: exploring the link between Helicobacter pylori and gastric cancer. Aliment Pharmacol Ther. 1999;13 Suppl 1:3-11.

37. Kusters JG, van Vliet AH, Kuipers EJ. Pathogenesis of Helicobacter pylori infection. Clin Microbiol Rev. 2006;19(3):449-90.

38. Engels EA. Epidemiologic perspectives on immunosuppressed populations and the immunosurveillance and immunocontainment of cancer. Am J Transplant. 2019;19(12):3223-32.

39. Huang JQ, Zheng GF, Sumanac K, Irvine EJ*, et al.* Meta-analysis of the relationship between cagA seropositivity and gastric cancer. Gastroenterology. 2003;125(6):1636-44.

40. Matos JI, de Sousa HA, Marcos-Pinto R, Dinis-Ribeiro M. *Helicobacter pylori* CagA and VacA genotypes and gastric phenotype: a meta-analysis. Eur J Gastroenterol Hepatol. 2013;25(12):1431-41.

41. Munoz N, Bosch FX, de Sanjose S, Herrero R*, et al.* Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med. 2003;348(6):518-27.

42. Cho LY, Yang JJ, Ko KP, Park B*, et al.* Coinfection of hepatitis B and C viruses and risk of hepatocellular carcinoma: systematic review and meta-analysis. Int J Cancer. 2011;128(1):176-84.

43. International Agency for Research on Agency. List of classifications by cancer sites with sufficient or limited evidence in humans, IARC Monographs Volumes 1–130 2021 [Available from: https://monographs.iarc.who.int/wp-content/uploads/2019/07/Classifications_by_cancer_site.pdf].

44. World Cancer Research Fund. Diet, nutrition, physical activity and stomach cancer: https://www.wcrf.org/wp-content/uploads/2021/02/stomach-cancer-report.pdf. 2018.

45. Volesky KD, El-Zein M, Franco EL, Brenner DR*, et al.* Cancers attributable to infections in Canada. Prev Med. 2019;122:109-17.

46. Plummer M, Franceschi S, Vignat J, Forman D*, et al.* Global burden of gastric cancer attributable to *Helicobacter pylori*. Int J Cancer. 2015;136(2):487-90.

47. D'Souza G, Kreimer AR, Viscidi R, Pawlita M*, et al.* Case-control study of human papillomavirus and oropharyngeal cancer. N Engl J Med. 2007;356(19):1944-56.

48. Hansson BG, Rosenquist K, Antonsson A, Wennerberg J*, et al.* Strong association between infection with human papillomavirus and oral and oropharyngeal squamous cell carcinoma: a population-based case-control study in southern Sweden. Acta Otolaryngol. 2005;125(12):1337-44.

49. Pintos J, Black MJ, Sadeghi N, Ghadirian P*, et al.* Human papillomavirus infection and oral cancer: a case-control study in Montreal, Canada. Oral Oncol. 2008;44(3):242-50.

50. Zhu H, Shen Z, Luo H, Zhang W*, et al.* Chlamydia Trachomatis Infection-Associated Risk of Cervical Cancer: A Meta-Analysis. Medicine (Baltimore). 2016;95(13):e3077.

51. Plummer M, Herrero R, Franceschi S, Meijer CJ*, et al.* Smoking and cervical cancer: pooled analysis of the IARC multi-centric case--control study. Cancer Causes Control. 2003;14(9):805-14.

52. Liu G, Sharma M, Tan N, Barnabas RV. HIV-positive women have higher risk of human papilloma virus infection, precancerous lesions, and cervical cancer. AIDS. 2018;32(6):795-808.

53. Madeleine MM, Finch JL, Lynch CF, Goodman MT*, et al.* HPV-related cancers after solid organ transplantation in the United States. Am J Transplant. 2013;13(12):3202-9.

54. de Martel C, Ferlay J, Franceschi S, Vignat J*, et al.* Global burden of cancer attributable to infections in 2008: a review and synthetic analysis. The Lancet. 2012;13:607-15.

55. Parkin D, Pisani P, Munoz N, Ferlay J. The global health burden of infection associated cancers. In: Weiss R, Beral V, Newton R, editors. Infections and human cancer. 33. New York: Cold Spring Harbor Laboratory Press; 1999. p. 5-33.

56. Plummer M, de Martel C, Vigant J, Ferlay J*, et al.* Global burden of cancer attributable to infections in 2012: a synthetic analysis. The Lancet. 2016;4(9):e609-e16.

57. de Martel C, Georges D, Bray F, Ferlay J*, et al.* Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. Lancet Glob Health. 2020;8(2):e180-e90.

58. Antonsson A, Wilson LF, Kendall BJ, Bain CJ*, et al.* Cancers in Australia in 2010 attributable to infectious agents. Aust N Z J Public Health. 2015;39(5):446-51.

59. Azevedo ESG, de Moura L, Curado MP, Gomes Fda S*, et al.* The Fraction of Cancer Attributable to Ways of Life, Infections, Occupation, and Environmental Agents in Brazil in 2020. PLoS One. 2016;11(2):e0148761.

60. Xiang W, Shi JF, Li P, Wang JB*, et al.* Estimation of cancer cases and deaths attributable to infection in China. Cancer Causes Control. 2011;22(8):1153-61.

61. Islami F, Chen W, Yu XQ, Lortet-Tieulent J*, et al.* Cancer deaths and cases attributable to lifestyle factors and infections in China, 2013. Ann Oncol. 2017;28(10):2567-74.

62. Shield KD, Marant Micallef C, de Martel C, Heard I*, et al.* New cancer cases in France in 2015 attributable to infectious agents: a systematic review and meta-analysis. Eur J Epidemiol. 2018;33(3):263-74.

63. International Agency for Research on Cancer. Attributable Causes of Cancer in France in the Year 2000. Lyon, France: World Health Organization; 2007.

64. De Flora S, Crocetti E, Bonanni P, Ferro A*, et al.* Incidence of infection-associated cancers in Italy and prevention strategies. Epidemiol Prev. 2015;39(4 Suppl 1):14-20.

65. Shin A, Park S, Shin HR, Park EH*, et al.* Population attributable fraction of infection-related cancers in Korea. Ann Oncol. 2011;22(6):1435-42.

66. van Lier EA, van Kranen HJ, van Vliet JA, Rahamat-Langendoen JC. Estimated number of new cancer cases attributable to infection in the Netherlands in 2003. Cancer Lett. 2008;272(2):226-31.

67. Parkin DM. Cancers attributable to infection in the UK in 2010. Br J Cancer. 2011;105 Suppl 2:S49-56.

68. Odutola M, Jedy-Agba EE, Dareng EO, Oga EA*, et al.* Burden of Cancers Attributable to Infectious Agents in Nigeria: 2012-2014. Front Oncol. 2016;6:216.

69. Levin ML. The occurrence of lung cancer in man. Acta Unio Int Contra Cancrum. 1953;9(3):531-41.

70. Brenner DR, Poirier AE, Walter SD, King WD*, et al.* Estimating the current and future cancer burden in Canada: methodological framework of the Canadian population attributable risk of cancer (ComPARe) study. BMJ Open. 2018;8(7):e022378.

71. Walter SD. Local estimates of population attributable risk. J Clin Epidemiol. 2010;63(1):85-93.

72. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. Am J Epidemiol. 1974;99(5):325-32.

73. de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. Int J Cancer. 2017.

74. Makarova-Rusher OV, Altekruse SF, McNeel TS, Ulahannan S*, et al.* Population attributable fractions of risk factors for hepatocellular carcinoma in the United States. Cancer. 2016;122(11):1757-65.

75. Engel LS, Chow WH, Vaughan TL, Gammon MD*, et al.* Population attributable risks of esophageal and gastric cancers. J Natl Cancer Inst. 2003;95(18):1404-13.

76. Morgenstern H, Bursic ES. A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. J Community Health. 1982;7(4):292-309.

77. Helicobacter Cancer Collaborative Group. Gastric cancer and *Helicobacter pylori:* a combined analysis of 12 case control studies nested within prospective cohorts. Gut. 2001;49(3):347-53.

78. Mesri EA, Cesarman E, Boshoff C. Kaposi's sarcoma and its associated herpesvirus. Nat Rev Cancer. 2010;10(10):707-19.

79. De Flora S, Bonanni P. The prevention of infection-associated cancers. Carcinogenesis. 2011;32(6):787-95.

80. Falade-Nwulia O, Suarez-Cuervo C, Nelson DR, Fried MW*, et al.* Oral Direct-Acting Agent Therapy for Hepatitis C Virus Infection: A Systematic Review. Ann Intern Med. 2017;166(9):637-48.

81. Kohli A, Shaffer A, Sherman A, Kottilil S. Treatment of hepatitis C: a systematic review. JAMA. 2014;312(6):631-40.

82. Lingala S, Ghany MG. Natural History of Hepatitis C. Gastroenterol Clin North Am. 2015;44(4):717-34.

83. Li X, Gao L, Li H, Gao J*, et al.* Human papillomavirus infection and laryngeal cancer risk: a systematic review and meta-analysis. J Infect Dis. 2013;207(3):479-88.

84. Torrente MC, Rodrigo JP, Haigentz M, Jr., Dikkers FG*, et al.* Human papillomavirus infections in laryngeal cancer. Head Neck. 2011;33(4):581-6.

85. International Agency for Research on Cancer. Epstein-Barr Virus and Kaposi's Sarcoma Herpesvirus/Human Herpesvirus 8. World Health Organization; 1997.

86. International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Human Papillomavirus. 2007.

87. Bruni L B-RL, Albero G, Serrano B, Mena M, Gómez D, Muñoz J, Bosch FX, de Sanjosé S. Human Papillomavirus and Related Diseases in USA. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre); 2017.

88. Bruni L B-RL, Albero G, Serrano B, Mena M, Gómez D, Muñoz J, Bosch FX, de Sanjosé S. Human Papillomavirus and Related Diseases in Canada. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre); 2017.

89. Stein AP, Saha S, Yu M, Kimple RJ*, et al.* Prevalence of human papillomavirus in oropharyngeal squamous cell carcinoma in the United States across time. Chem Res Toxicol. 2014;27(4):462-9.

90. Trubnikov M, Yan P, Archibald C. Estimated prevalence of Hepatitis C Virus infection in Canada, 2011. Can Commun Dis Rep. 2014;40(19):429-36.

91. Rotermann M, Langlois K, Andonov A, M T. Seroprevalence of hepatitis B and C virus infections: Results from the 2007 to 2009 and 2009 to 2011 Canadian Health Measures Survey. Statistics Canada. Health Reports. 2013;82-003-X.

92. Statistics Canada. Canadian Health Measures Survey (CHMS) [Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5071].

93. National Notifiable Diseases Surveillance System. Hepatitis B, Chronic, 2012 Case Definition: Centers for Disease Control and Prevention; 2012 [Available from: https://wwwn.cdc.gov/nndss/conditions/hepatitis-b-chronic/case-definition/2012/].

94. Centers for Disease Control and Prevention. NHANES 2007-2008 Laboratory Data. In: Department of Health and Human Services, editor. 2009.

95. Centers for Disease Control and Prevention. NHANES 2009-2010 Laboratory Data. In: Department of Health and Human Services, editor. 2011.

96. El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology. 2012;142(6):1264-73 e1.

97. Statistics Canada. Canadian Health Measures Survey (CHMS) Data User Guide: Cycle 1. 2010.

98. Public Health Agency of Canada. List of nationally notifiable diseases: Government of Canada; 2018 [Available from: http://diseases.canada.ca/notifiable/diseases-list].

99. Remis R. Modelling the incidence of prevalence of hepatitis C infection and its sequelae in Canada, 2007. Public Health Agency of Canada; 2010.

100. Cheung J, Goodman KJ, Girgis S, Bailey R*, et al.* Disease manifestations of Helicobacter pylori infection in Arctic Canada: using epidemiology to address community concerns. BMJ Open. 2014;4(1):e003689.

101. Sethi A, Chaudhuri M, Kelly L, Hopman W. Prevalence of *Helicobacter pylori* in a First Nations population in northwestern Ontario. Can Fam Physician. 2013;59(4):e182-7.

102. Naja F, Kreiger N, Sullivan T. *Helicobacter pylori* infection in Ontario: prevalence and risk factors. Can J Gastroenterol. 2007;21(8):501-6.

103. Centers for Disease Control and Prevention. NHANES 1999-2000 Laboratory Data. In: Department of Health and Human Services, editor. 2001.

104. Monteiro L, de Mascarel A, Sarrasqueta AM, Bergey B*, et al.* Diagnosis of *Helicobacter pylori* infection: noninvasive methods compared to invasive methods and evaluation of two new tests. Am J Gastroenterol. 2001;96(2):353-8.

105. Franco EL. Measurement errors in epidemiological studies of human papillaomvirus and cervical cancer. In: Muñoz NB, F.X.; Meheus, A., editor. The Epidemiology of Human Papillomavirus and Cervical Cancer. 119. Lyon: International Agency for Reseach on Cancer; 1992.

106. Gonzalez CA, Megraud F, Buissonniere A, Lujan Barroso L*, et al.* Helicobacter pylori infection assessed by ELISA and by immunoblot and noncardia gastric cancer risk in a prospective study: the Eurgast-EPIC project. Ann Oncol. 2012;23(5):1320-4.

107. Mitchell H, English DR, Elliott F, Gengos M*, et al.* Immunoblotting using multiple antigens is essential to demonstrate the true risk of Helicobacter pylori infection for gastric cancer. Aliment Pharmacol Ther. 2008;28(7):903-10.

108. Peleteiro B, Lunet N, Barros R, La Vecchia C*, et al.* Factors contributing to the underestimation of *Helicobacter pylori*-associated gastric cancer risk in a high-prevalence population. Cancer Causes Control. 2010;21(8):1257-64.

109. Altekruse SF, Devesa SS, Dickie LA, McGlynn KA*, et al.* Histological classification of liver and intrahepatic bile duct cancers in SEER registries. J Registry Manag. 2011;38(4):201-5.

110. Nyaga VN, Arbyn M, Aerts M. Metaprop: a Stata command to perform meta-analysis of binomial data. Arch Public Health. 2014;72(1):39.

111. Harris R BM, Deeks J, Harbord R, Altman D, Sterne J. metan: fixed- and random-effects meta-analysis. Stata Journal. 2008;8(1):3-28.

112. R: A language and environment for statistical computing. R Foundation for Statistical Computing [Internet]. 2017. Available from: https://www.R-project.org/.

113. Brenner DR, Poirier AE, Walter SD, King WD*, et al.* Estimating the current and future cancer burden in Canada: methodological framework of the Canadian population attributable risk of cancer (ComPARe) study. BMJ Open. 2018;8(7).

114. Brenner DR, Friedenreich CM, Ruan Y, Poirier AE*, et al.* The burden of cancer attributable to modifiable risk factors in Canada: Methodological overview. To be submitted to Preventive Medicine. 2019.

115. Bourke B, Ceponis P, Chiba N, Czinn S*, et al.* Canadian Helicobacter Study Group Consensus Conference: Update on the approach to *Helicobacter pylor*i infection in children and adolescents--an evidence-based evaluation. Can J Gastroenterol. 2005;19(7):399-408.

116. Fallone CA, Chiba N, van Zanten SV, Fischbach L*, et al.* The Toronto Consensus for the Treatment of *Helicobacter pylori* Infection in Adults. Gastroenterology. 2016;151(1):51-69 e14.

117. Ndiaye C, Mena M, Alemany L, Arbyn M*, et al.* HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: a systematic review and meta-analysis. Lancet Oncol. 2014;15(12):1319-31.

118. Mehanna H, Beech T, Nicholson T, El-Hariry I*, et al.* Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer--systematic review and meta-analysis of trends by time and region. Head Neck. 2013;35(5):747-55.

119. Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2015. Toronto, ON: Canadian Cancer Society, .

120. Gulley ML, Tang W. Laboratory assays for Epstein-Barr virus-related disease. J Mol Diagn. 2008;10(4):279-92.

121. Goncalves DU, Proietti FA, Ribas JG, Araujo MG*, et al.* Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases. Clin Microbiol Rev. 2010;23(3):577-89.

122. Parsonnet J, Hansen S, Rodriguez L, Gelb AB*, et al. Helicobacter pylori* infection and gastric lymphoma. N Engl J Med. 1994;330(18):1267-71.

123. Anderson LA, Pfeiffer R, Warren JL, Landgren O*, et al.* Hematopoietic malignancies associated with viral and alcoholic hepatitis. Cancer Epidemiol Biomarkers Prev. 2008;17(11):3069-75.

124. Schweitzer A, Horn J, Mikolajczyk RT, Krause G*, et al.* Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. Lancet. 2015;386(10003):1546-55.

125. Hyams KC. Risks of chronicity following acute hepatitis B virus infection: a review. Clin Infect Dis. 1995;20(4):992-1000.

126. International Agency for Research on Cancer. Hepatitis Viruses. Lyon, France; 1994.

127. Micallef JM, Kaldor JM, Dore GJ. Spontaneous viral clearance following acute hepatitis C infection: a systematic review of longitudinal studies. J Viral Hepat. 2006;13(1):34-41.

128. Pel P.
De ziekten van de maag met het oog op de behoeften der geneeskundige praktijk geschet st. In: Bohn EF, editor. 3rd ed. Haarlem, the Netherlands: De Erven; 1913.

129. Eusebi LH, Zagari RM, Bazzoli F. Epidemiology of *Helicobacter pylori* infection. Helicobacter. 2014;19 Suppl 1:1-5.

130. Hansen S, Vollset SE, Derakhshan MH, Fyfe V*, et al.* Two distinct aetiologies of cardia cancer; evidence from premorbid serological markers of gastric atrophy and *Helicobacter pylori* status. Gut. 2007;56(7):918-25.

131. Knekt P, Teppo L, Aromaa A, Rissanen H*, et al. Helicobacter pylori* IgA and IgG antibodies, serum pepsinogen I and the risk of gastric cancer: changes in the risk with extended follow-up period. Int J Cancer. 2006;119(3):702-5.

132. Nomura AM, Lee J, Stemmermann GN, Nomura RY*, et al. Helicobacter pylori* CagA seropositivity and gastric carcinoma risk in a Japanese American population. J Infect Dis. 2002;186(8):1138-44.

133. Siman JH, Engstrand L, Berglund G, Forsgren A*, et al. Helicobacter pylori* and CagA seropositivity and its association with gastric and oesophageal carcinoma. Scand J Gastroenterol. 2007;42(8):933-40.

134. Canadian Cancer Society. MALT lymphoma 2017 [Available from: http://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/malt-lymphoma/?region=bc].

135. Zullo A, Hassan C, Cristofari F, Andriani A*, et al.* Effects of *Helicobacter pylori* eradication on early stage gastric mucosa-associated lymphoid tissue lymphoma. Clin Gastroenterol Hepatol. 2010;8(2):105-10.

136. Epstein MA, Achong BG, Barr YM. Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma. Lancet. 1964;1(7335):702-3.

137. Ambinder RF, Cesarman E. Clinical and pathological aspects of EBV and KSHV infection. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge2007.

138. Crawford DH, Swerdlow AJ, Higgins C, McAulay K*, et al.* Sexual history and Epstein-Barr virus infection. J Infect Dis. 2002;186(6):731-6.

139. Khan G, Hashim MJ. Global burden of deaths from Epstein-Barr virus attributable malignancies 1990-2010. Infect Agent Cancer. 2014;9(1):38.

140. Mbulaiteye SM, Pullarkat ST, Nathwani BN, Weiss LM*, et al.* Epstein-Barr virus patterns in US Burkitt lymphoma tumors from the SEER residual tissue repository during 1979-2009. APMIS. 2014;122(1):5-15.

141. Dean AG SK, Soe MM. Simple proportion. OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version OpenEpicom, updated 2013/04/06 [Internet]. 2013.

142. Metgud RS, Doshi JJ, Gaurkhede S, Dongre R*, et al.* Extranodal NK/T-cell lymphoma, nasal type (angiocentric T-cell lymphoma): A review about the terminology. J Oral Maxillofac Pathol. 2011;15(1):96-100.

143. Park S, Ko YH. Epstein-Barr virus-associated T/natural killer-cell lymphoproliferative disorders. J Dermatol. 2014;41(1):29-39.

144. Canadian Cancer Society. Types of Hodgkin lymphoma [Available from: http://www.cancer.ca/en/cancer-information/cancer-type/hodgkin-lymphoma/hodgkin-lymphoma/types-of-hodgkin-lymphoma/?region=on].

145. Lee JH, Kim Y, Choi JW, Kim YS. Prevalence and prognostic significance of Epstein-Barr virus infection in classical Hodgkin's lymphoma: a meta-analysis. Arch Med Res. 2014;45(5):417-31.

146. Keegan TH, Glaser SL, Clarke CA, Gulley ML*, et al.* Epstein-Barr virus as a marker of survival after Hodgkin's lymphoma: a population-based study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2005;23(30):7604-13.

147. Chang ET, Zheng T, Lennette ET, Weir EG*, et al.* Heterogeneity of risk factors and antibody profiles in epstein-barr virus genome-positive and -negative hodgkin lymphoma. 2004;189(12):2271-81.

148. Vasef MA, Ubaidat MA, Khalidi HS, Almasri NM*, et al.* Association between Epstein-Barr virus and classic Hodgkin lymphoma in Jordan: a comparative study with Epstein-Barr virus-associated Hodgkin lymphoma in North America. South Med J. 2004;97(3):273-7.

149. Elenitoba-Johnson KS, Medeiros LJ, Khorsand J, King TC. P53 expression in Reed-Sternberg cells does not correlate with gene mutations in Hodgkin's disease. Am J Clin Pathol. 1996;106(6):728-38.

150. Shah KM, Young LS. Epstein-Barr virus and carcinogenesis: beyond Burkitt's lymphoma. Clin Microbiol Infect. 2009;15(11):982-8.

151. Dogan S, Hedberg ML, Ferris RL, Rath TJ*, et al.* Human papillomavirus and Epstein-Barr virus in nasopharyngeal carcinoma in a low-incidence population. Head Neck. 2014;36(4):511-6.

152. Shi W, Pataki I, MacMillan C, Pintilie M*, et al.* Molecular pathology parameters in human nasopharyngeal carcinoma. Cancer. 2002;94(7):1997-2006.

153. International Agency for Research on Cancer. Human Papillomaviruses. Lyon, France; 1995.

154. Schiffman M, Doorbar J, Wentzensen N, de Sanjose S*, et al.* Carcinogenic human papillomavirus infection. Nat Rev Dis Primers. 2016;2:16086.

155. Bosch FX, Broker TR, Forman D, Moscicki AB*, et al.* Comprehensive control of human papillomavirus infections and related diseases. Vaccine. 2013;31 Suppl 6:G1-31.

156. Ho GY, Burk RD, Klein S, Kadish AS*, et al.* Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. J Natl Cancer Inst. 1995;87(18):1365-71.

157. Moscicki AB, Schiffman M, Burchell A, Albero G*, et al.* Updating the natural history of human papillomavirus and anogenital cancers. Vaccine. 2012;30 Suppl 5:F24-33.

158. Forman D, de Martel C, Lacey CJ, Soerjomataram I*, et al.* Global burden of human papillomavirus and related diseases. Vaccine. 2012;30 Suppl 5:F12-23.

159. Chung JH, Sanford E, Johnson A, Klempner SJ*, et al.* Comprehensive genomic profiling of anal squamous cell carcinoma reveals distinct genomically defined classes. Ann Oncol. 2016;27(7):1336-41.

160. Alemany L, Saunier M, Alvarado-Cabrero I, Quiros B*, et al.* Human papillomavirus DNA prevalence and type distribution in anal carcinomas worldwide. Int J Cancer. 2015;136(1):98-107.

161. Ouhoummane N, Steben M, Coutlee F, Vuong T*, et al.* Squamous anal cancer: patient characteristics and HPV type distribution. Cancer Epidemiol. 2013;37(6):807-12.

162. Meyer JE, Panico VJ, Marconato HM, Sherr DL*, et al.* HIV positivity but not HPV/p16 status is associated with higher recurrence rate in anal cancer. J Gastrointest Cancer. 2013;44(4):450-5.

163. Herfs M, Longuespee R, Quick CM, Roncarati P*, et al.* Proteomic signatures reveal a dualistic and clinically relevant classification of anal canal carcinoma. J Pathol. 2017;241(4):522-33.

164. Steinau M, Unger ER, Hernandez BY, Goodman MT*, et al.* Human papillomavirus prevalence in invasive anal cancers in the United States before vaccine introduction. J Low Genit Tract Dis. 2013;17(4):397-403.

165. Alemany L, Cubilla A, Halec G, Kasamatsu E*, et al.* Role of Human Papillomavirus in Penile Carcinomas Worldwide. Eur Urol. 2016;69(5):953-61.

166. McDaniel AS, Hovelson DH, Cani AK, Liu CJ*, et al.* Genomic Profiling of Penile Squamous Cell Carcinoma Reveals New Opportunities for Targeted Therapy. Cancer Res. 2015;75(24):5219-27.

167. Hernandez BY, Goodman MT, Unger ER, Steinau M*, et al.* Human papillomavirus genotype prevalence in invasive penile cancers from a registry-based United States population. Front Oncol. 2014;4:9.

168. Daling JR, Madeleine MM, Johnson LG, Schwartz SM*, et al.* Penile cancer: importance of circumcision, human papillomavirus and smoking in in situ and invasive disease. Int J Cancer. 2005;116(4):606-16.

169. Rubin MA, Kleter B, Zhou M, Ayala G*, et al.* Detection and typing of human papillomavirus DNA in penile carcinoma: evidence for multiple independent pathways of penile carcinogenesis. Am J Pathol. 2001;159(4):1211-8.

170. Cupp MR, Malek RS, Goellner JR, Smith TF*, et al.* The detection of human papillomavirus deoxyribonucleic acid in intraepithelial, in situ, verrucous and invasive carcinoma of the penis. J Urol. 1995;154(3):1024-9.

171. Sinno AK, Saraiya M, Thompson TD, Hernandez BY*, et al.* Human papillomavirus genotype prevalence in invasive vaginal cancer from a registry-based population. Obstet Gynecol. 2014;123(4):817-21.

172. Daling JR, Madeleine MM, Schwartz SM, Shera KA*, et al.* A population-based study of squamous cell vaginal cancer: HPV and cofactors. Gynecol Oncol. 2002;84(2):263-70.

173. Gargano JW, Wilkinson EJ, Unger ER, Steinau M*, et al.* Prevalence of human papillomavirus types in invasive vulvar cancers and vulvar intraepithelial neoplasia 3 in the United States before vaccine introduction. J Low Genit Tract Dis. 2012;16(4):471-9.

174. de Koning MN, Quint WG, Pirog EC. Prevalence of mucosal and cutaneous human papillomaviruses in different histologic subtypes of vulvar carcinoma. Mod Pathol. 2008;21(3):334-44.

175. Al-Ghamdi A, Freedman D, Miller D, Poh C*, et al.* Vulvar squamous cell carcinoma in young women: a clinicopathologic study of 21 cases. Gynecol Oncol. 2002;84(1):94-101.

176. Kim YT, Thomas NF, Kessis TD, Wilkinson EJ*, et al.* p53 mutations and clonality in vulvar carcinomas and squamous hyperplasias: evidence suggesting that squamous hyperplasias do not serve as direct precursors of human papillomavirus-negative vulvar carcinomas. Hum Pathol. 1996;27(4):389-95.

177. Bishop JA, Ma XJ, Wang H, Luo Y*, et al.* Detection of transcriptionally active high-risk HPV in patients with head and neck squamous cell carcinoma as visualized by a novel E6/E7 mRNA in situ hybridization method. Am J Surg Pathol. 2012;36(12):1874-82.

178. Rietbergen MM, Leemans CR, Bloemena E, Heideman DA*, et al.* Increasing prevalence rates of HPV attributable oropharyngeal squamous cell carcinomas in the Netherlands as assessed by a validated test algorithm. Int J Cancer. 2013;132(7):1565-71.

179. Biron VL, Kostiuk M, Isaac A, Puttagunta L*, et al.* Detection of human papillomavirus type 16 in oropharyngeal squamous cell carcinoma using droplet digital polymerase chain reaction. Cancer. 2016;122(10):1544-51.

180. Hooper JE, Hebert JF, Schilling A, Gross ND*, et al.* Hybrid Capture 2 is as effective as PCR testing for high-risk human papillomavirus in head and neck cancers. Appl Immunohistochem Mol Morphol. 2015;23(4):266-72.

181. Isayeva T, Xu J, Dai Q, Whitley AC*, et al.* African Americans with oropharyngeal carcinoma have significantly poorer outcomes despite similar rates of human papillomavirus-mediated carcinogenesis. Hum Pathol. 2014;45(2):310-9.

182. Nichols AC, Dhaliwal SS, Palma DA, Basmaji J*, et al.* Does HPV type affect outcome in oropharyngeal cancer? J Otolaryngol Head Neck Surg. 2013;42:9.

183. Lingen MW, Xiao W, Schmitt A, Jiang B*, et al.* Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas. Oral Oncol. 2013;49(1):1-8.

184. Walline HM, Komarck C, McHugh JB, Byrd SA*, et al.* High-risk human papillomavirus detection in oropharyngeal, nasopharyngeal, and oral cavity cancers: comparison of multiple methods. JAMA Otolaryngol Head Neck Surg. 2013;139(12):1320-7.

185. Jordan RC, Lingen MW, Perez-Ordonez B, He X*, et al.* Validation of methods for oropharyngeal cancer HPV status determination in US cooperative group trials. Am J Surg Pathol. 2012;36(7):945-54.

186. Stephen JK, Chen KM, Shah V, Havard S*, et al.* Human papillomavirus outcomes in an access-to-care laryngeal cancer cohort. Otolaryngol Head Neck Surg. 2012;146(5):730-8.

187. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY*, et al.* Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol. 2011;29(32):4294-301.

188. Schlecht NF, Brandwein-Gensler M, Nuovo GJ, Li M*, et al.* A comparison of clinically utilized human papillomavirus detection methods in head and neck cancer. Mod Pathol. 2011;24(10):1295-305.

189. Agoston ES, Robinson SJ, Mehra KK, Birch C*, et al.* Polymerase chain reaction detection of HPV in squamous carcinoma of the oropharynx. Am J Clin Pathol. 2010;134(1):36-41.

190. Jo S, Juhasz A, Zhang K, Ruel C*, et al.* Human papillomavirus infection as a prognostic factor in oropharyngeal squamous cell carcinomas treated in a prospective phase II clinical trial. Anticancer Res. 2009;29(5):1467-74.

191. Settle K, Posner MR, Schumaker LM, Tan M*, et al.* Racial survival disparity in head and neck cancer results from low prevalence of human papillomavirus infection in black oropharyngeal cancer patients. Cancer Prev Res (Phila). 2009;2(9):776-81.

192. Tezal M, Sullivan Nasca M, Stoler DL, Melendy T*, et al.* Chronic periodontitis-human papillomavirus synergy in base of tongue cancers. Arch Otolaryngol Head Neck Surg. 2009;135(4):391-6.

193. Cohen MA, Basha SR, Reichenbach DK, Robertson E*, et al.* Increased viral load correlates with improved survival in HPV-16-associated tonsil carcinoma patients. Acta Otolaryngol. 2008;128(5):583-9.

194. Liang XH, Lewis J, Foote R, Smith D*, et al.* Prevalence and significance of human papillomavirus in oral tongue cancer: the Mayo Clinic experience. J Oral Maxillofac Surg. 2008;66(9):1875-80.

195. Worden FP, Kumar B, Lee JS, Wolf GT*, et al.* Chemoselection as a strategy for organ preservation in advanced oropharynx cancer: response and survival positively associated with HPV16 copy number. J Clin Oncol. 2008;26(19):3138-46.

196. Zhao M, Rosenbaum E, Carvalho AL, Koch W*, et al.* Feasibility of quantitative PCR-based saliva rinse screening of HPV for head and neck cancer. Int J Cancer. 2005;117(4):605-10.

197. Ha PK, Pai SI, Westra WH, Gillison ML*, et al.* Real-time quantitative PCR demonstrates low prevalence of human papillomavirus type 16 in premalignant and malignant lesions of the oral cavity. Clin Cancer Res. 2002;8(5):1203-9.

198. Strome SE, Savva A, Brissett AE, Gostout BS*, et al.* Squamous cell carcinoma of the tonsils: a molecular analysis of HPV associations. Clin Cancer Res. 2002;8(4):1093-100.

199. Robey RC, Bower M. Facing up to the ongoing challenge of Kaposi's sarcoma. Curr Opin Infect Dis. 2015;28(1):31-40.

200. O'Brien SF, Goldman M, Scalia V, Yi QL*, et al.* The epidemiology of human T-cell lymphotropic virus types I and II in Canadian blood donors. Transfus Med. 2013;23(5):358-66.

201. Matutes E. Adult T-cell leukaemia/lymphoma. J Clin Pathol. 2007;60(12):1373-7.

202. Harris Insights & Analytics. National Cancer Opinion Survey Prepared for American Society of Clinical Oncology 2017 [Available from: https://www.asco.org/sites/new-www.asco.org/files/content-files/research-and-progress/documents/2017-ASCO-National-Cancer-Opinion-Survey-Results.pdf].

203. American Cancer Society. Viruses that Can Lead to Cancer. 2021 [Available from: https://www.cancer.org/cancer/cancer-causes/infectious-agents/infections-that-can-lead-to-cancer/viruses.html].

204. Chen XZ, Chen H, Castro FA, Hu JK*, et al.* Epstein-Barr virus infection and gastric cancer: a systematic review. Medicine (Baltimore). 2015;94(20):e792.

205. Bae JM, Kim EH. Epstein-Barr Virus and Gastric Cancer Risk: A Meta-analysis With Meta-regression of Case-control Studies. J Prev Med Public Health. 2016;49(2):97-107.

206. Hwang J, Suh CH, Won Kim K, Kim HS*, et al.* The Incidence of Epstein-Barr Virus-Positive Diffuse Large B-Cell Lymphoma: A Systematic Review and Meta-Analysis. Cancers (Basel). 2021;13(8).

207. Centers for Disease Control and Prevention. About Opisthorchis. 2020 [Available from: https://www.cdc.gov/parasites/opisthorchis/faqs.html].

208. Centers for Disease Control and Prevention. Parasites - Clonorchis. 2021 [Available from: https://www.cdc.gov/parasites/clonorchis/index.html].

209. Centers for Disease Control and Prevention. About Schistosomiasis. 2021 [Available from: https://www.cdc.gov/parasites/schistosomiasis/gen_info/faqs.html].

210. Hassan WM, Bakry MS, Hassan HM, Alfaar AS. Incidence of orbital, conjunctival and lacrimal gland malignant tumors in USA from Surveillance, Epidemiology and End Results, 1973-2009. Int J Ophthalmol. 2016;9(12):1808-13.

211. Dal Maso L, Franceschi S. Hepatitis C virus and risk of lymphoma and other lymphoid neoplasms: a meta-analysis of epidemiologic studies. Cancer Epidemiol Biomarkers Prev. 2006;15(11):2078-85.

212. Morton LM, Slager SL, Cerhan JR, Wang SS*, et al.* Etiologic heterogeneity among non-Hodgkin lymphoma subtypes: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr. 2014;2014(48):130-44.

213. Zhou Y, Zhao Y, Li B, Huang J*, et al.* Hepatitis viruses infection and risk of intrahepatic cholangiocarcinoma: evidence from a meta-analysis. BMC Cancer. 2012;12:289.

214. Li M, Li J, Li P, Li H*, et al.* Hepatitis B virus infection increases the risk of cholangiocarcinoma: a meta-analysis and systematic review. J Gastroenterol Hepatol. 2012;27(10):1561-8.

215. Palmer WC, Patel T. Are common factors involved in the pathogenesis of primary liver cancers? A meta-analysis of risk factors for intrahepatic cholangiocarcinoma. J Hepatol. 2012;57(1):69-76.

216. Li H, Hu B, Zhou ZQ, Guan J*, et al.* Hepatitis C virus infection and the risk of intrahepatic cholangiocarcinoma and extrahepatic cholangiocarcinoma: evidence from a systematic review and meta-analysis of 16 case-control studies. World J Surg Oncol. 2015;13:161.

217. Ok CY, Papathomas TG, Medeiros LJ, Young KH. EBV-positive diffuse large B-cell lymphoma of the elderly. Blood. 2013;122(3):328-40.

218. Teras LR, Rollison DE, Pawlita M, Michel A*, et al.* Epstein-Barr virus and risk of non-Hodgkin lymphoma in the cancer prevention study-II and a meta-analysis of serologic studies. Int J Cancer. 2015;136(1):108-16.

219. American Cancer Society. Types of B-cell Lymphoma. 2019 [Available from: https://www.cancer.org/cancer/non-hodgkin-lymphoma/about/b-cell-lymphoma.html].

220. American Cancer Society. Cancer Stat Facts: Stomach Cancer. 2021 [Available from: https://seer.cancer.gov/statfacts/html/stomach.html].

221. Rokkas T, Sechopoulos P, Pistiolas D, Kothonas F*, et al.* The relationship of *Helicobacter pylori* infection and colon neoplasia, on the basis of meta-analysis. Eur J Gastroenterol Hepatol. 2013;25(11):1286-94.

222. Islami F, Kamangar F. *Helicobacter pylori* and esophageal cancer risk: a meta-analysis. Cancer Prev Res (Phila). 2008;1(5):329-38.

223. Zhuo X, Zhang Y, Wang Y, Zhuo W*, et al. Helicobacter pylori* infection and oesophageal cancer risk: association studies via evidence-based meta-analyses. Clin Oncol (R Coll Radiol). 2008;20(10):757-62.

224. Nie S, Chen T, Yang X, Huai P*, et al.* Association of *Helicobacter pylori* infection with esophageal adenocarcinoma and squamous cell carcinoma: a meta-analysis. Dis Esophagus. 2014;27(7):645-53.

225. Gao H, Li L, Zhang C, Tu J*, et al.* Systematic Review with Meta-analysis: Association of *Helicobacter pylori* Infection with Esophageal Cancer. Gastroenterol Res Pract. 2019;2019:1953497.

226. Xie FJ, Zhang YP, Zheng QQ, Jin HC*, et al. Helicobacter pylori* infection and esophageal cancer risk: an updated meta-analysis. World J Gastroenterol. 2013;19(36):6098-107.

227. de Martel C, Shiels MS, Franceschi S, Simard EP*, et al.* Cancers attributable to infections among adults with HIV in the United States. AIDS. 2015;29(16):2173-81.

228. International Agency for Research on Cancer. Human Immunodeficiency Viruses and Human T-Cell Lymphotropic Viruses. Lyon, France; 1996.

229. Bruni L, Albero G, Serrano B, Mena M*, et al.* Human Papillomavirus and Related Diseases in United States of America. 2021.

230. Johnson CL, Paulose-Ram R, Ogden CL, Carroll MD*, et al.* National health and nutrition examination survey: analytic guidelines, 1999-2010. Vital Health Stat 2. 2013(161):1-24.

231. De Silva AP, De Livera AM, Lee KJ, Moreno-Betancur M*, et al.* Multiple imputation methods for handling missing values in longitudinal studies with sampling weights: Comparison of methods implemented in Stata. Biom J. 2021;63(2):354-71.

232. National Center for Health Statistics (NCHS). Module 4: Variance Estimation [Available from: https://wwwn.cdc.gov/nchs/nhanes/tutorials/module4.aspx].

233. Clopper C, Pearson E. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 1934;26(4):404-13.

234. Dean A, Sullivan K, Soe M. OpenEpi: Open Source Epidemiologic Statistics for Public Health. 2013.

235. United States Department of Health and Human Services, Centers for Disease Control and Prevention, National Cancer Institute. National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER*Stat Database: NPCR and SEER Incidence - U.S. Cancer Statistics 2001-2017 Public Use Research Database. Accessed at www.cdc.gov/cancer/uscs/public-use. 2020.

236. Shiels MS, Koritzinsky EH, Clarke CA, Suneja G*, et al.* Prevalence of HIV Infection among U.S. Hodgkin lymphoma cases. Cancer Epidemiol Biomarkers Prev. 2014;23(2):274-81.

237. Shiels MS, Pfeiffer RM, Hall HI, Li J*, et al.* Proportions of Kaposi sarcoma, selected non-Hodgkin lymphomas, and cervical cancer in the United States occurring in persons with AIDS, 1980-2007. JAMA. 2011;305(14):1450-9.

238. Denniston MM, Jiles RB, Drobeniuc J, Klevens RM*, et al.* Chronic hepatitis C virus infection in the United States, National Health and Nutrition Examination Survey 2003 to 2010. Ann Intern Med. 2014;160(5):293-300.

239. Roberts H, Kruszon-Moran D, Ly KN, Hughes E*, et al.* Prevalence of chronic hepatitis B virus (HBV) infection in U.S. households: National Health and Nutrition Examination Survey (NHANES), 1988-2012. Hepatology. 2016;63(2):388-97.

240. Roberts H, Ly KN, Yin S, Hughes E*, et al.* Prevalence of HBV Infection, Vaccine-Induced Immunity, and Susceptibility Among At-Risk Populations: US Households, 2013-2018. Hepatology. 2021;74(5):2353-65.

241. Watts T, Lauver D, Sethi AK, Snedden T*, et al.* Hepatitis C virus infections among people aged 15-44, United States, 2009-2018. Public Health Nurs. 2021;38(2):167-75.

242. Edlin BR, Eckhardt BJ, Shu MA, Holmberg SD*, et al.* Toward a more accurate estimate of the prevalence of hepatitis C in the United States. Hepatology. 2015;62(5):1353-63.

243. Hall EW, Rosenberg ES, Sullivan PS. Estimates of state-level chronic hepatitis C virus infection, stratified by race and sex, United States, 2010. BMC Infect Dis. 2018;18(1):224.

244. Rajbhandari R, Chung RT. Treatment of Hepatitis B: A Concise Review. Clin Transl Gastroenterol. 2016;7(9):e190.

245. Balfour HH, Jr., Schmeling DO, Grimm-Geris JM. The promise of a prophylactic Epstein-Barr virus vaccine. Pediatr Res. 2020;87(2):345-52.

246. Cohen JI. Vaccine Development for Epstein-Barr Virus. Adv Exp Med Biol. 2018;1045:477-93.

247. Frisch M, Fenger C, van den Brule AJ, Sorensen P*, et al.* Variants of squamous cell carcinoma of the anal canal and perianal skin and their relation to human papillomaviruses. Cancer Res. 1999;59(3):753-7.

248. Guan P, Howell-Jones R, Li N, Bruni L*, et al.* Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. Int J Cancer. 2012;131(10):2349-59.

249. Petrick JL, Yang B, Altekruse SF, Van Dyke AL*, et al.* Risk factors for intrahepatic and extrahepatic cholangiocarcinoma in the United States: A population-based study in SEER-Medicare. PLoS One. 2017;12(10):e0186643.

250. Shaib YH, El-Serag HB, Nooka AK, Thomas M*, et al.* Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: a hospital-based case-control study. Am J Gastroenterol. 2007;102(5):1016-21.

251. Surveillance, Epidemiology, and End Results Program. Primary effusion lymphoma. 2020 [Available from: https://seer.cancer.gov/seertools/hemelymph/51f6cf57e3e27c3994bd5378/].

252. Persson C, Jia Y, Pettersson H, Dillner J*, et al. H. pylori* Seropositivity before Age 40 and Subsequent Risk of Stomach Cancer: A Glimpse of the True Relationship? PLoS One. 2011;6(3):e17404.

253. Parsonnet J, Samloff IM, Nelson LM, Orentreich N*, et al. Helicobacter pylori*, pepsinogen, and risk for gastric adenocarcinoma. Cancer Epidemiol Biomarkers Prev. 1993;2(5):461-6.

254. de Martel C, Llosa AE, Farr SM, Friedman GD*, et al.* Helicobacter pylori infection and the risk of development of esophageal adenocarcinoma. J Infect Dis. 2005;191(5):761-7.

255. El-Omar EM, Rabkin CS, Gammon MD, Vaughan TL*, et al.* Increased risk of noncardia gastric cancer associated with proinflammatory cytokine gene polymorphisms. Gastroenterology. 2003;124(5):1193-201.

256. Wu AH, Crabtree JE, Bernstein L, Hawtin P*, et al.* Role of *Helicobacter pylori* CagA+ strains and risk of adenocarcinoma of the stomach and esophagus. Int J Cancer. 2003;103(6):815-21.

257. Fruh M, Zhou W, Zhai R, Su L*, et al.* Polymorphisms of inflammatory and metalloproteinase genes, Helicobacter pylori infection and the risk of oesophageal adenocarcinoma. Br J Cancer. 2008;98(4):689-92.

258. Davila JA, Morgan RO, Shaib Y, McGlynn KA*, et al.* Diabetes increases the risk of hepatocellular carcinoma in the United States: a population based case control study. Gut. 2005;54(4):533-9.

259. Hassan MM, Hwang LY, Hatten CJ, Swaim M*, et al.* Risk factors for hepatocellular carcinoma: synergism of alcohol with viral hepatitis and diabetes mellitus. Hepatology. 2002;36(5):1206-13.

260. Nomura A, Stemmermann GN, Chyou PH, Tabor E. Hepatitis B and C virus serologies among Japanese Americans with hepatocellular carcinoma. J Infect Dis. 1996;173(6):1474-6.

261. Di Bisceglie AM, Order SE, Klein JL, Waggoner JG*, et al.* The role of chronic viral hepatitis in hepatocellular carcinoma in the United States. Am J Gastroenterol. 1991;86(3):335-8.

262. Choi J, Ghoz HM, Peeraphatdit T, Baichoo E*, et al.* Aspirin use and the risk of cholangiocarcinoma. Hepatology. 2016;64(3):785-96.

263. Shaib YH, El-Serag HB, Davila JA, Morgan R*, et al.* Risk factors of intrahepatic cholangiocarcinoma in the United States: a case-control study. Gastroenterology. 2005;128(3):620-6.

264. Ognjanovic S, Yuan JM, Chaptman AK, Fan Y*, et al.* Genetic polymorphisms in the cytokine genes and risk of hepatocellular carcinoma in low-risk non-Asians of USA. Carcinogenesis. 2009;30(5):758-62.

265. Hassan MM, Spitz MR, Thomas MB, Curley SA*, et al.* The association of family history of liver cancer with hepatocellular carcinoma: a case-control study in the United States. J Hepatol. 2009;50(2):334-41.

266. Welzel TM, Graubard BI, El-Serag HB, Shaib YH*, et al.* Risk factors for intrahepatic and extrahepatic cholangiocarcinoma in the United States: a population-based case-control study. Clin Gastroenterol Hepatol. 2007;5(10):1221-8.

267. Naeini YB, Wu A, O'Malley DP. Aggressive B-cell lymphomas: frequency, immunophenotype, and genetics in a reference laboratory population. Ann Diagn Pathol. 2016;25:7-14.

268. Zhao LH, Liu X, Yan HX, Li WY*, et al.* Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. Nat Commun. 2016;7:12992.

269. Morton LM, Sampson JN, Cerhan JR, Turner JJ*, et al.* Rationale and Design of the International Lymphoma Epidemiology Consortium (InterLymph) Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr. 2014;2014(48):1-14.

270. Mbulaiteye SM, Morton LM, Sampson JN, Chang ET*, et al.* Medical history, lifestyle, family history, and occupational risk factors for sporadic Burkitt lymphoma/leukemia: the Interlymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr. 2014;2014(48):106-14.

271. Slager SL, Benavente Y, Blair A, Vermeulen R*, et al.* Medical history, lifestyle, family history, and occupational risk factors for chronic lymphocytic leukemia/small lymphocytic lymphoma: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr. 2014;2014(48):41-51.

272. Cerhan JR, Kricker A, Paltiel O, Flowers CR*, et al.* Medical history, lifestyle, family history, and occupational risk factors for diffuse large B-cell lymphoma: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr. 2014;2014(48):15-25.

273. Vajdic CM, Landgren O, McMaster ML, Slager SL*, et al.* Medical history, lifestyle, family history, and occupational risk factors for lymphoplasmacytic lymphoma/Waldenstrom's macroglobulinemia: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr. 2014;2014(48):87-97.

274. Bracci PM, Benavente Y, Turner JJ, Paltiel O*, et al.* Medical history, lifestyle, family history, and occupational risk factors for marginal zone lymphoma: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr. 2014;2014(48):52-65.

275. Peleteiro B, Bastos A, Ferro A, Lunet N. Prevalence of *Helicobacter pylori* infection worldwide: a systematic review of studies with national coverage. Dig Dis Sci. 2014;59(8):1698-709.

276. Yang L, Kartsonaki C, Yao P, de Martel C*, et al.* The relative and attributable risks of cardia and non-cardia gastric cancer associated with Helicobacter pylori infection in China: a case-cohort study. Lancet Public Health. 2021;6(12):e888-e96.

277. Colquhoun A, Arnold M, Ferlay J, Goodman KJ*, et al.* Global patterns of cardia and non-cardia gastric cancer incidence in 2012. Gut. 2015;64(12):1881-8.

278. Cavaleiro-Pinto M, Peleteiro B, Lunet N, Barros H. Helicobacter pylori infection and gastric cardia cancer: systematic review and meta-analysis. Cancer Causes Control. 2011;22(3):375-87.

279. Lochhead P, El-Omar EM. Helicobacter pylori infection and gastric cancer. Best Pract Res Clin Gastroenterol. 2007;21(2):281-97.

280. Zucca E, Arcaini L, Buske C, Johnson PW*, et al.* Marginal zone lymphomas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2020;31(1):17-29.

281. Ladetto M, Dreyling M. EHA Endorsement of ESMO Clinical Practice Guidelines for Marginal Zone Lymphomas. Hemasphere. 2020;4(2):e351.

282. Jung K, Kim DH, Seo HI, Gong EJ*, et al.* Efficacy of eradication therapy in *Helicobacter pylori*-negative gastric mucosa-associated lymphoid tissue lymphoma: A meta-analysis. Helicobacter. 2021;26(2):e12774.

283. Zhang Y. Epidemiology of esophageal cancer. World J Gastroenterol. 2013;19(34):5598-606.

284. Patel N, Benipal B. Incidence of Esophageal Cancer in the United States from 2001-2015: A United States Cancer Statistics Analysis of 50 States. Cureus. 2018;10(12):e3709.

285. Whiteman DC, Parmar P, Fahey P, Moore SP*, et al.* Association of *Helicobacter pylori* infection with reduced risk for esophageal cancer is independent of environmental and genetic modifiers. Gastroenterology. 2010;139(1):73-83; quiz e11-2.

286. El-Omar EM, Oien K, El-Nujumi A, Gillen D*, et al.* Helicobacter pylori infection and chronic gastric acid hyposecretion. Gastroenterology. 1997;113(1):15-24.

287. Polyzos SA, Zeglinas C, Artemaki F, Doulberis M*, et al.* Helicobacter pylori infection and esophageal adenocarcinoma: a review and a personal view. Ann Gastroenterol. 2018;31(1):8-13.

288. Gall A, Fero J, McCoy C, Claywell BC*, et al.* Bacterial Composition of the Human Upper Gastrointestinal Tract Microbiome Is Dynamic and Associated with Genomic Instability in a Barrett's Esophagus Cohort. PLoS One. 2015;10(6):e0129055.

289. Holleczek B, Schottker B, Brenner H. Helicobacter pylori infection, chronic atrophic gastritis and risk of stomach and esophagus cancer: Results from the prospective population-based ESTHER cohort study. Int J Cancer. 2020;146(10):2773-83.

290. Vicari JJ, Peek RM, Falk GW, Goldblum JR*, et al.* The seroprevalence of cagA-positive *Helicobacter pylori* strains in the spectrum of gastroesophageal reflux disease. Gastroenterology. 1998;115(1):50-7.

291. Chow WH, Blaser MJ, Blot WJ, Gammon MD*, et al.* An inverse relation between cagA+ strains of *Helicobacter pylori* infection and risk of esophageal and gastric cardia adenocarcinoma. Cancer Res. 1998;58(4):588-90.

292. Weston AP, Badr AS, Topalovski M, Cherian R*, et al.* Prospective evaluation of the prevalence of gastric *Helicobacter pylori* infection in patients with GERD, Barrett's esophagus, Barrett's dysplasia, and Barrett's adenocarcinoma. Am J Gastroenterol. 2000;95(2):387-94.

293. Anandasabapathy S, Jhamb J, Davila M, Wei C*, et al.* Clinical and endoscopic factors predict higher pathologic grades of Barrett dysplasia. Cancer. 2007;109(4):668-74.

294. Oberg S, Peters JH, Nigro JJ, Theisen J*, et al. Helicobacter pylori* is not associated with the manifestations of gastroesophageal reflux disease. Arch Surg. 1999;134(7):722-6.

295. Peek RM, Jr., Vaezi MF, Falk GW, Goldblum JR*, et al.* Role of *Helicobacter pylori* cagA(+) strains and specific host immune responses on the development of premalignant and malignant lesions in the gastric cardia. Int J Cancer. 1999;82(4):520-4.

296. Richter J, John K, Staiger AM, Rosenwald A*, et al.* Epstein-Barr virus status of sporadic Burkitt lymphoma is associated with patient age and mutational features. Br J Haematol. 2021;196(3):681-9.

297. Dupont Harwood C, Eriksen PRG, Clasen-Linde E, Jensen JS*, et al.* Clinicopathologic characteristics of Burkitt lymphoma of the head and neck in a non-endemic region-a Danish nationwide study. Acta Oto Laryngologica. 2021.

298. Kasprzak A, Spachacz R, Wachowiak J, Stefanska K*, et al.* Epstein-Barr virus (EBV) infection in B-cell non-Hodgkin's lymphomas in children: Virus latency and its correlation with CD21 and CD23 molecules. 2007;45(3):169-79.

299. Karajannis MA, Hummel M, Oschlies I, Anagnostopoulos I*, et al.* Epstein-Barr virus infection and viral gene expression in pediatric non-Hodgkin lymphomas. In: Stathopoulou FT, editor. Genome and Proteome in Oncology. New York: Nova Biomedical; 2005. p. 297-325.

300. Teitell MA, Lones MA, Perkins SL, Sanger WG*, et al.* TCL1 expression and Epstein-Barr virus status in pediatric Burkitt lymphoma.124(4):569-75.

301. Haralambieva E, Schuuring E, Rosati S, Van Noesel C*, et al.* Interphase Fluorescence In Situ Hybridization for Detection of 8q24/MYC Breakpoints on Routine Histologic Sections: Validation in Burkitt Lymphomas from Three Geographic Regions. Genes Chromosomes and Cancer. 2004;40(1):10-8.

302. Lenze D, Leoncini L, Hummel M, Volinia S*, et al.* The different epidemiologic subtypes of Burkitt lymphoma share a homogenous micro RNA profile distinct from diffuse large B-cell lymphoma.25(12):1869-76.

303. Hansen PB, Penkowa M, Kirk O, Skinhoj P*, et al.* Human immunodeficiency virus-associated malignant lymphoma in eastern Denmark diagnosed from 1990-1996: clinical features, histopathology, and association with Epstein-Barr virus and human herpesvirus-8. Eur J Haematol. 2000;64(6):368-75.

304. Davi F, Delecluse HJ, Guiet P, Gabarre J*, et al.* Burkitt-like lymphomas in AIDS patients: characterization within a series of 103 human immunodeficiency virus-associated non-Hodgkin's lymphomas. Burkitt's Lymphoma Study Group. J Clin Oncol. 1998;16(12):3788-95.

305. Spina M, Tirelli U, Zagonel V, Gloghini A*, et al.* Burkitt's lymphoma in adults with and without human immunodeficiency virus infection: a single-institution clinicopathologic study of 75 patients. Cancer. 1998;82(4):766-74.

306. Glaser SL, Clarke CA, Gulley ML, Craig FE*, et al.* Population-based patterns of human immunodeficiency virus-related Hodgkin lymphoma in the Greater San Francisco Bay Area, 1988-1998. Cancer. 2003;98(2):300-9.

307. Thompson LD, Fisher SI, Chu WS, Nelson A*, et al.* HIV-associated Hodgkin lymphoma: a clinicopathologic and immunophenotypic study of 45 cases. Am J Clin Pathol. 2004;121(5):727-38.

308. Besson C, Lancar R, Prevot S, Brice P*, et al.* High Risk Features Contrast With Favorable Outcomes in HIV-associated Hodgkin Lymphoma in the Modern cART Era, ANRS CO16 LYMPHOVIR Cohort. Clin Infect Dis. 2015;61(9):1469-75.

309. Hentrich M, Berger M, Wyen C, Siehl J*, et al.* Stage-adapted treatment of HIV-associated Hodgkin lymphoma: results of a prospective multicenter study. J Clin Oncol. 2012;30(33):4117-23.

310. Carbone A, Gloghini A, Larocca LM, Antinori A*, et al.* Human immunodeficiency virus-associated Hodgkin's disease derives from post-germinal center B cells. Blood. 1999;93(7):2319-26.

311. Tirelli U, Errante D, Dolcetti R, Gloghini A*, et al.* Hodgkin's disease and human immunodeficiency virus infection: clinicopathologic and virologic features of 114 patients from the Italian Cooperative Group on AIDS and Tumors. J Clin Oncol. 1995;13(7):1758-67.

312. Linabery AM, Erhardt EB, Richardson MR, Ambinder RF*, et al.* Family history of cancer and risk of pediatric and adolescent Hodgkin lymphoma: A Children's Oncology Group study. 2015;137(9):2163-74.

313. Siddon A, Lozovatsky L, Mohamed A, Hudnall SD. Human herpesvirus 6 positive Reed-Sternberg cells in nodular sclerosis Hodgkin lymphoma. British Journal of Haematology. 2012;158(5):635-43.

314. Glaser SL, Gulley ML, Clarke CA, Keegan TH*, et al.* Racial/ethnic variation in EBV-positive classical Hodgkin lymphoma in California populations.123(7):1499-507.

315. Heller KN, Arrey F, Steinherz P, Portlock C*, et al.* Patients with Epstein Barr virus-positive lymphomas have decreased CD4(+) T-cell responses to the viral nuclear antigen 1. 2008;123(12):2824-31.

316. Andriko JW, Aguilera NS, edkar MA, Abbondanzo SL. Childhood Hodgkin's disease in the United States: An analysis of histologic subtypes and association with Epstein-Barr virus. 1997;10(4):366-71.

317. Razzouk BI, Gan YJ, Mendonca C, Jenkins JJ*, et al.* Epstein-Barr virus in pediatric Hodgkin disease: Age and histiotype are more predictive than geographic region. Medical and Pediatric Oncology. 1997;28(4):248-54.

318. Lin AY, Kingma DW, Lennette ET, Fears TR*, et al.* Epstein-Barr virus and familial Hodgkin's disease. Blood. 1996;88(8):3160-5.

319. Mertens R, Granzen B, Lassay L, Gademann G*, et al.* Nasopharyngeal carcinoma in childhood and adolescence: concept and preliminary results of the cooperative GPOH study NPC-91. Gesellschaft fur Padiatrische Onkologie und Hamatologie. Cancer. 1997;80(5):951-9.

320. Polychronopoulou S, Kostaridou S, Panagiotou JP, Stefanaki K*, et al.* Nasopharyngeal carcinoma in childhood and adolescence: A single institution's experience with treatment modalities during the last 15 years.21(5):393-402.

321. Verma N, Patel S, Osborn V, McBride S*, et al.* Prognostic significance of human papillomavirus and Epstein-Bar virus in nasopharyngeal carcinoma. Head Neck. 2020;42(9):2364-74.

322. Jiang W, Chamberlain PD, Garden AS, Kim BY*, et al.* Prognostic value of p16 expression in Epstein-Barr virus-positive nasopharyngeal carcinomas. Head Neck. 2016;38 Suppl 1:E1459-66.

323. Lin Z, Khong B, Kwok S, Cao H*, et al.* Human papillomavirus 16 detected in nasopharyngeal carcinomas in white Americans but not in endemic Southern Chinese patients. Head Neck. 2014;36(5):709-14.

324. Stenmark MH, McHugh JB, Schipper M, Walline HM*, et al.* Nonendemic HPV-positive nasopharyngeal carcinoma: association with poor prognosis. Int J Radiat Oncol Biol Phys. 2014;88(3):580-8.

325. Wilson DD, Crandley EF, Sim A, Stelow EB*, et al.* Prognostic significance of p16 and its relationship with human papillomavirus in pharyngeal squamous cell carcinomas. JAMA Otolaryngol Head Neck Surg. 2014;140(7):647-53.

326. Singhi AD, Califano J, Westra WH. High-risk human papillomavirus in nasopharyngeal carcinoma. Head Neck. 2012;34(2):213-8.

327. Haverkos BM, Pan Z, Gru AA, Freud AG*, et al.* Extranodal NK/T Cell Lymphoma, Nasal Type (ENKTL-NT): An Update on Epidemiology, Clinical Presentation, and Natural History in North American and European Cases. Curr Hematol Malig Rep. 2016;11(6):514-27.

328. Li S, Feng X, Li T, Zhang S*, et al.* Extranodal NK/T-cell lymphoma, nasal type: a report of 73 cases at MD Anderson Cancer Center. Am J Surg Pathol. 2013;37(1):14-23.

329. Chan JKC, Quintanilla-Martinez L, Ferry JA, Peh SC. Extranodal NK/T-cell lymphoma, nasal type. In: Swerdlow SH, Campo E, Harris NK, Jaffe ES, Pileri SA, Stein H, et al., editors. World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon: IARC Press; 20008. p. 285-8.

330. Bourbon E, Maucort-Boulch D, Fontaine J, Mauduit C*, et al.* Clinicopathological features and survival in EBV-positive diffuse large B-cell lymphoma not otherwise specified. Blood Adv. 2021;5(16):3227-39.

331. Keane C, Tobin J, Gunawardana J, Francis S*, et al.* The tumour microenvironment is immuno-tolerogenic and a principal determinant of patient outcome in EBV-positive diffuse large B-cell lymphoma. Eur J Haematol. 2019;103(3):200-7.

332. Tracy SI, Habermann TM, Feldman AL, Maurer MJ*, et al.* Outcomes among North American patients with diffuse large B-cell lymphoma are independent of tumor Epstein-Barr virus positivity or immunosuppression. Haematologica. 2018;103(2):297-303.

333. Petrella T, Copie-Bergman C, Briere J, Delarue R*, et al.* BCL2 expression but not MYC and BCL2 coexpression predicts survival in elderly patients with diffuse large B-cell lymphoma independently of cell of origin in the phase 3 LNH03-6B trial. Ann Oncol. 2017;28(5):1042-9.

334. Tisi MC, Cupelli E, Santangelo R, Maiolo E*, et al.* Whole blood EBV-DNA predicts outcome in diffuse large B-cell lymphoma. Leuk Lymphoma. 2016;57(3):628-34.

335. Ziarkiewicz M, Wolosz D, Dzieciatkowski T, Wilczek E*, et al.* Epstein-Barr Virus-Positive Diffuse Large B cell Lymphoma in the Experience of a Tertiary Medical Center in Poland. Arch Immunol Ther Exp (Warsz). 2016;64(2):159-69.

336. Morton LM, Kim CJ, Weiss LM, Bhatia K*, et al.* Molecular characteristics of diffuse large B-cell lymphoma in human immunodeficiency virus-infected and -uninfected patients in the pre-highly active antiretroviral therapy and pre-rituximab era. Leuk Lymphoma. 2014;55(3):551-7.

337. Ok CY, Li L, Xu-Monette ZY, Visco C*, et al.* Prevalence and clinical implications of epstein-barr virus infection in de novo diffuse large B-cell lymphoma in Western countries. Clin Cancer Res. 2014;20(9):2338-49.

338. Slack GW, Steidl C, Sehn LH, Gascoyne RD. CD30 expression in de novo diffuse large B-cell lymphoma: a population-based study from British Columbia. Br J Haematol. 2014;167(5):608-17.

339. Hofscheier A, Ponciano A, Bonzheim I, Adam P*, et al.* Geographic variation in the prevalence of Epstein-Barr virus-positive diffuse large B-cell lymphoma of the elderly: a comparative analysis of a Mexican and a German population. Mod Pathol. 2011;24(8):1046-54.

340. Gibson SE, Hsi ED. Epstein-Barr virus-positive B-cell lymphoma of the elderly at a United States tertiary medical center: an uncommon aggressive lymphoma with a nongerminal center B-cell phenotype. Hum Pathol. 2009;40(5):653-61.

341. Hoeller S, Tzankov A, Pileri SA, Went P*, et al.* Epstein-Barr virus–positive diffuse large B-cell lymphoma in elderly patients is rare in Western populations. Human Pathology. 2010;41(3):352-7.

342. d'Amore F, Johansen P, Houmand A, Weisenburger DD*, et al.* Epstein-Barr virus genome in non-Hodgkin's lymphomas occurring in immunocompetent patients: highest prevalence in nonlymphoblastic T-cell lymphoma and correlation with a poor prognosis. Danish Lymphoma Study Group, LYFO. Blood. 1996;87(3):1045-55.

343. Ramos JC, Sparano JA, Chadburn A, Reid EG*, et al.* Impact of Myc in HIV-associated non-Hodgkin lymphomas treated with EPOCH and outcomes with vorinostat (AMC-075 trial). Blood. 2020;136(11):1284-97.

344. Chao C, Silverberg MJ, Martinez-Maza O, Chi M*, et al.* Epstein-Barr virus infection and expression of B-cell oncogenic markers in HIV-related diffuse large B-cell Lymphoma. Clin Cancer Res. 2012;18(17):4702-12.

345. Chadburn A, Chiu A, Lee JY, Chen X*, et al.* Immunophenotypic analysis of AIDS-related diffuse large B-cell lymphoma and clinical implications in patients from AIDS Malignancies Consortium clinical trials 010 and 034. J Clin Oncol. 2009;27(30):5039-48.

346. Vaghefi P, Martin A, Prevot S, Charlotte F*, et al.* Genomic imbalances in AIDS-related lymphomas: relation with tumoral Epstein-Barr virus status. AIDS. 2006;20(18):2285-91.

347. Burke AP, Yen TS, Shekitka KM, Sobin LH. Lymphoepithelial carcinoma of the stomach with Epstein-Barr virus demonstrated by polymerase chain reaction. Mod Pathol. 1990;3(3):377-80.

348. Shibata D, Weiss LM. Epstein-Barr virus-associated gastric adenocarcinoma. Am J Pathol. 1992;140(4):769-74.

349. Murphy G, Pfeiffer R, Camargo MC, Rabkin CS. Meta-analysis shows that prevalence of Epstein-Barr virus-positive gastric cancer differs based on sex and anatomic location. Gastroenterology. 2009;137(3):824-33.

350. Li S, Du H, Wang Z, Zhou L*, et al.* Meta-analysis of the relationship between Epstein-Barr virus infection and clinicopathological features of patients with gastric carcinoma. Sci China Life Sci. 2010;53(4):524-30.

351. Sousa H, Pinto-Correia AL, Medeiros R, Dinis-Ribeiro M. Epstein-Barr virus is associated with gastric carcinoma: the question is what is the significance? World J Gastroenterol. 2008;14(27):4347-51.

352. Camargo MC, Murphy G, Koriyama C, Pfeiffer RM*, et al.* Determinants of Epstein-Barr virus-positive gastric cancer: an international pooled analysis. Br J Cancer. 2011;105(1):38-43.

353. Lee JH, Kim SH, Han SH, An JS*, et al.* Clinicopathological and molecular characteristics of Epstein-Barr virus-associated gastric carcinoma: a meta-analysis. J Gastroenterol Hepatol. 2009;24(3):354-65.

354. Tavakoli A, Monavari SH, Solaymani Mohammadi F, Kiani SJ*, et al.* Association between Epstein-Barr virus infection and gastric cancer: a systematic review and meta-analysis. BMC Cancer. 2020;20(1):493.

355. Kim TS, da Silva E, Coit DG, Tang LH. Intratumoral Immune Response to Gastric Cancer Varies by Molecular and Histologic Subtype. Am J Surg Pathol. 2019;43(6):851-60.

356. Ma C, Patel K, Singhi AD, Ren B*, et al.* Programmed Death-Ligand 1 Expression Is Common in Gastric Cancer Associated With Epstein-Barr Virus or Microsatellite Instability. Am J Surg Pathol. 2016;40(11):1496-506.

357. Truong CD, Feng W, Li W, Khoury T*, et al.* Characteristics of Epstein-Barr virus-associated gastric cancer: a study of 235 cases at a comprehensive cancer center in U.S.A. J Exp Clin Cancer Res. 2009;28:14.

358. Grogg KL, Lohse CM, Pankratz VS, Halling KC*, et al.* Lymphocyte-rich gastric cancer: associations with Epstein-Barr virus, microsatellite instability, histology, and survival. Mod Pathol. 2003;16(7):641-51.

359. Vo QN, Geradts J, Gulley ML, Boudreau DA*, et al.* Epstein-Barr virus in gastric adenocarcinomas: association with ethnicity and CDKN2A promoter methylation. J Clin Pathol. 2002;55(9):669-75.

360. Shibata D, Hawes D, Stemmermann GN, Weiss LM. Epstein-Barr virus-associated gastric adenocarcinoma among Japanese Americans in Hawaii. Cancer Epidemiol Biomarkers Prev. 1993;2(3):213-7.

361. Kreuter A, Potthoff A, Brockmeyer NH, Gambichler T*, et al.* Anal carcinoma in human immunodeficiency virus-positive men: results of a prospective study from Germany. Br J Dermatol. 2010;162(6):1269-77.

362. Arana R, Flejou JF, Si-Mohamed A, Bauer P*, et al.* Clinicopathological and virological characteristics of superficially invasive squamous-cell carcinoma of the anus. Colorectal Dis. 2015;17(11):965-72.

363. Zhu X, Jamshed S, Zou J, Azar A*, et al.* Molecular and immunophenotypic characterization of anal squamous cell carcinoma reveals distinct clinicopathologic groups associated with HPV and TP53 mutation status. Mod Pathol. 2021;34(5):1017-30.

364. De Vuyst H, Clifford GM, Nascimento MC, Madeleine MM*, et al.* Prevalence and type distribution of human papillomavirus in carcinoma and intraepithelial neoplasia of the vulva, vagina and anus: a meta-analysis. Int J Cancer. 2009;124(7):1626-36.

365. Kolitz E, Lucas E, Hosler GA, Kim J*, et al.* Human Papillomavirus Positive and Negative Vulvar Squamous Cell Carcinoma Are Biologically but Not Clinically Distinct. J Invest Dermatol. 2021.

366. Lewis JS, Jr., Mirabello L, Liu P, Wang X*, et al.* Oropharyngeal Squamous Cell Carcinoma Morphology and Subtypes by Human Papillomavirus Type and by 16 Lineages and Sublineages. Head Neck Pathol. 2021;15(4):1089-98.

367. Mazul AL, Rodriguez-Ormaza N, Taylor JM, Desai DD*, et al.* Prognostic significance of non-HPV16 genotypes in oropharyngeal squamous cell carcinoma. Oral Oncol. 2016;61:98-103.

368. Zandberg DP, Liu S, Goloubeva OG, Schumaker LM*, et al.* Emergence of HPV16-positive oropharyngeal cancer in Black patients over time: University of Maryland 1992-2007. Cancer Prev Res (Phila). 2015;8(1):12-9.

369. Kingma DW, Allen RA, Moore W, Caughron SK*, et al.* HPV genotype distribution in oral and oropharyngeal squamous cell carcinoma using seven in vitro amplification assays. Anticancer Res. 2010;30(12):5099-104.

370. American Cancer Society. Risk Factors and Causes of Childhood Cancer 2021 [Available from: https://www.cancer.org/cancer/cancer-in-children/risk-factors-and-causes.html].

371. Cancer Research UK. Risks and causes of cancer in children 2020 [Available from: https://www.cancerresearchuk.org/about-cancer/childrens-cancer/risks-causes].

372. Kyu HH, Stein CE, Boschi Pinto C, Rakovac I*, et al.* Causes of death among children aged 5-14 years in the WHO European Region: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Child Adolesc Health. 2018;2(5):321-37.

373. Cunningham RM, Walton MA, Carter PM. The Major Causes of Death in Children and Adolescents in the United States. N Engl J Med. 2018;379(25):2468-75.

374. Sung H, Ferlay J, Siegel RL, Laversanne M*, et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.

375. Molyneux EM, Rochford R, Griffin B, Newton R*, et al.* Burkitt's lymphoma. Lancet. 2012;379(9822):1234-44.

376. Bray F, Haugen M, Moger TA, Tretli S*, et al.* Age-incidence curves of nasopharyngeal carcinoma worldwide: bimodality in low-risk populations and aetiologic implications. Cancer Epidemiol Biomarkers Prev. 2008;17(9):2356-65.

377. Zhou L, Deng Y, Li N, Zheng Y*, et al.* Global, regional, and national burden of Hodgkin lymphoma from 1990 to 2017: estimates from the 2017 Global Burden of Disease study. J Hematol Oncol. 2019;12(1):107.

378. Moher D, Shamseer L, Clarke M, Ghersi D*, et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4:1.

379. Wilberger AC, Liang X. Primary Nonanaplastic Peripheral Natural Killer/T-Cell Lymphoma in Pediatric Patients-An Unusual Distribution Pattern of Subtypes.22(2):128-36.

380. Delecluse HJ, Feederle R, O'Sullivan B, Taniere P. Epstein Barr virus-associated tumours: an update for the attention of the working pathologist. J Clin Pathol. 2007;60(12):1358-64.

381. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan — a web and mobile app for systematic reviews. Systematic Reviews. 2016.

382. The EndNote Team. EndNote. EndNote X9 ed. Philadelphia, PA: Clarivate; 2013.

383. Munn Z, Moola S, Lisy K, Riitano D*, et al.* Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. Int J Evid Based Healthc. 2015;13(3):147-53.

384. Anagnostopoulos I, Hansmann ML, Franssila K, Harris M*, et al.* European Task Force on Lymphoma project on lymphocyte predominance Hodgkin disease: histologic and immunohistologic analysis of submitted cases reveals 2 types of Hodgkin disease with a nodular growth pattern and abundant lymphocytes. Blood. 2000;96(5):1889-99.

385. Carbone A, Gloghini A, Dotti G. EBV-associated lymphoproliferative disorders: classification and treatment. Oncologist. 2008;13(5):577-85.

386. Steliarova-Foucher E, Colombet M, Ries L, Hesseling P*, et al.* International Incidence of Childhood Cancer, Volume III (electronic version: http://iicc.iarc.fr/results/). Lyon, France: International Agency for Research on Cancer; 2017.

387. Brousset P, Rochaix P, Chittal S, Rubie H*, et al.* High incidence of Epstein-Barr virus detection in Hodgkin's disease and absence of detection in anaplastic large-cell lymphoma in children. Histopathology. 1993;23(2):189-91.

388. Karajannis MA, Hummel M, Oschlies I, Anagnostopoulos I*, et al.* Epstein-Barr virus infection in Western European pediatric non-Hodgkin lymphomas [2].102(12):4244.

389. Mbulaiteye SM, Pullarkat ST, Nathwani BN, Weiss LM*, et al.* Epstein-Barr virus patterns in US Burkitt lymphoma tumors from the SEER residual tissue repository during 1979-2009.122(1):5-15.

390. Ambinder RF, Browning PJ, Lorenzana I, Leventhal BG*, et al.* Epstein-Barr virus and childhood Hodgkin's disease in Honduras and the United States. Blood. 1993;81(2):462-7.

391. Armstrong AA, Alex, er FE, Cartwright R*, et al.* Epstein-Barr virus and Hodgkin's disease: Further evidence for the three disease hypothesis. 1998;12(8):1272-6.

392. Bigenwald C, Galimard JE, Amorim S, Thieblemont C*, et al.* Hodgkin lymphoma in adolescents and young adults: Results from a single-tertiary-center prospective cohort of 349 patients. Haematologica. 2017;101:39.

393. Claviez A, Tiemann M, Luders H, Krams M*, et al.* Impact of latent Epstein-Barr virus infection on outcome in children and adolescents with Hodgkin's lymphoma. Journal of Clinical Oncology. 2005;23(18):4048-56.

394. Claviez A, Tiemann M, Peters J, Kreipe H*, et al.* The impact of EBV, proliferation rate, and Bcl-2 expression in Hodgkin's disease in childhood. Annals of Hematology. 1994;68(2):61-6.

395. Diepstra A, Van Imhoff GW, Karim-Kos HE, Van Den Berg A*, et al.* HLA class II expression by Hodgkin Reed-Sternberg cells is an independent prognostic factor in classical Hodgkin's lymphoma. Journal of Clinical Oncology. 2007;25(21):3101-8.

396. Dilly-Feldis M, Aladjidi N, Refait JK, Parrens M*, et al.* Expression of PD-1/PD-L1 in children's classical Hodgkin lymphomas. Pediatric Blood and Cancer. 2019;66.

397. Enblad G, vej K, Sundstrom C, Pallesen G*, et al.* Epstein-Barr virus distribution in Hodgkin's disease in an unselected Swedish population. 1999;38(4):425-9.

398. Englund A, Molin D, Enblad G, Karlen J*, et al.* The role of tumour-infiltrating eosinophils, mast cells and macrophages in Classical and Nodular Lymphocyte Predominant Hodgkin Lymphoma in children. European Journal of Haematology. 2016;97(5):430-8.

399. Flavell KJ, Biddulph JP, Powell JE, Parkes SE*, et al.* South Asian ethnicity and material deprivation increase the risk of Epstein-Barr virus infection in childhood Hodgkin's disease.85(3):350-6.

400. Foss HD, Herbst H, Oelmann E, Samol J*, et al.* Lymphotoxin, tumour necrosis factor and interleukin-6 gene transcripts are present in Hodgkin and Reed-Sternberg cells of most Hodgkin's disease cases. British Journal of Haematology. 1993;84(4):627-35.

401. Hamdi L, Creidy R, Boudjemaa S, Hendel-Chavez H*, et al.* Frequent altered distribution of peripheral B-lymphocyte subsets in pediatric and adolescent patients with classical Hodgkin lymphoma. Leukemia & Lymphoma. 2021;62(2):300-7.

402. Herbst H, Raff T, Stein H. Phenotypic modulation of Hodgkin and Reed-Sternberg cells by Epstein-Barr virus. Journal of Pathology. 1996;179(1):54-9.

403. Herling M, Rassidakis GZ, Medeiros LJ, Vassilakopoulos TP*, et al.* Expression of Epstein-Barr virus latent membrane protein-1 in Hodgkin and Reed-Sternberg cells of classical Hodgkin's lymphoma: Associations with presenting features, serum interleukin 10 levels, and clinical outcome. 2003;9(6):2114-20.

404. Hollander P, Kamper P, Smedby KE, Enblad G*, et al.* High proportions of PD-1+ and PD-L1+ leukocytes in classical Hodgkin lymphoma microenvironment are associated with inferior outcome. Blood advances. 2017;1(18):1427-39.

405. Horton TM, Sheehan AM, Lopez-Terrada D, Hutchison RE*, et al.* Analysis of NF- B pathway proteins in pediatric Hodgkin lymphoma: Correlations with EBV status and clinical outcome - A Children's Oncology Group study. Lymphoma. 2012;2012(341629).

406. Huppmann AR, Nicolae A, Slack GW, Pittaluga S*, et al.* EBV may be expressed in the LP cells of nodular lymphocyte-predominant hodgkin lymphoma (NLPHL) in both children and adults. 2014;38(3):316-24.

407. Jarrett RF, Krajewski AS, Angus B, Freel*, et al.* The Scotland and Newcastle epidemiological study of Hodgkin's disease: Impact of histopathological review and EBV status on incidence estimates. 2003;56(11):811-6.

408. Kaczorowski S, Kaczorowska M, Christensson B. Expression of EBV encoded latent membrane protein 1 (LMP-1) and bcl-2 protein in childhood and adult Hodgkin's disease: Application of microwave irradiation for antigen retrieval. Leukemia and Lymphoma. 1994;13(3):273-83.

409. Kanavaros P, Sakalidou A, Tzardi M, Darivianaki K*, et al.* Frequent detection of Epstein-Barr virus (EBV), EBER transcripts and latent membrane protein-1 (LMP-1) in tumor cells in Hodgkin's disease arising in childhood. Pathology, Research & Practice. 1994;190(11):1026-30.

410. Khan G, Norton AJ, Slavin G. Epstein-Barr virus in Hodgkin disease: Relation to age and subtype. Cancer. 1993;71(10):3124-9.

411. Klekawka T, Balwierz W. Relationship between epstein-barr virus latency in pediatric Hodgkin Lymphoma and immunological status of the host. Haematologica. 2013;2:25.

412. Kordek R, Jesionek-Kupnicka D, Biernat W, Wozniak L. Expression of the latent membrane protein of Epstein-Barr virus in Hodgkin's disease. Age and subtype distribution in Polish patients. 1996;27(1):15-20.

413. Lacroix A, Jaccard A, Rouzioux C, Piguet C*, et al.* HHV-6 and EBV DNA quantitation in lymph nodes of 86 patients with Hodgkin's lymphoma. Journal of Medical Virology. 2007;79(9):1349-56.

414. Panayiotides J, Kanavaros P, Protopapa E, Vlachonikolis J*, et al.* Morphologic differences between latent membrane protein-1 (LMP-1)-positive and negative tumour cells in Epstein-Barr virus (EBV)-related childhood Hodgkin's disease. A morphometric study. 1996;192(3):210-4.

415. Pavlovic A, Glavina Durdov M, Capkun V, Jakelic Pitesa J*, et al.* Classical Hodgkin Lymphoma with Positive Epstein-Barr Virus Status is Associated with More FOXP3 Regulatory T Cells. 2016;22:2340-6.

416. Santon A, Martin C, Manzanal AI, Preciado MV*, et al.* Paediatric Hodgkin's disease in Spain: Association with Epstein-Barr virus strains carrying latent membrane protein-1 oncogene deletions and high frequency of dual infections. British Journal of Haematology. 1998;103(1):129-36.

417. Trimeche M, Bonnet C, Korbi S, Boniver J*, et al.* Association between Epstein-Barr virus and Hodgkin's lymphoma in Belgium: A pathological and virological study.48(7):1323-31.

418. Weinreb M, Day PJR, Murray PG, Raafat F*, et al.* Epstein-Barr virus (EBV) and Hodgkin's disease in children: Incidence of EBV latent membrane protein in malignant cells. Journal of Pathology. 1992;168(4):365-9.

419. Weinreb M, Day PJR, Niggli F, Powell JE*, et al.* The role of Epstein-Barr virus in Hodgkin's disease from different geographical areas. 1996;74(1):27-31.

420. Venkateswaran L, Gan YJ, Sixbey JW, Santana VM. Epstein-Barr virus infection in salivary gland tumors in children and young adults. Cancer. 2000;89(2):463-6.

421. Mergan F, Jaubert F, Sauvat F, Hartmann O*, et al.* Inflammatory myofibroblastic tumor in children: clinical review with anaplastic lymphoma kinase, Epstein-Barr virus, and human herpesvirus 8 detection analysis. J Pediatr Surg. 2005;40(10):1581-6.

422. Grobner SN, Worst BC, Weischenfeldt J, Buchhalter I*, et al.* The landscape of genomic alterations across childhood cancers. Nature. 2018;555(7696):321-7.

423. Zhang J, Nichols KE, Downing JR. Germline Mutations in Predisposition Genes in Pediatric Cancer. N Engl J Med. 2016;374(14):1391.

424. International Agency for Research on Cancer. WHO classification of tumours of haematopoietic and lymphoid tissues. 4th ed. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, et al., editors 2008.

425. Cader FZ, Kearns P, Young L, Murray P*, et al.* The contribution of the Epstein-Barr virus to the pathogenesis of childhood lymphomas. Cancer Treat Rev. 2010;36(4):348-53.

426. Carbone A, Gloghini A. Epstein Barr Virus-Associated Hodgkin Lymphoma. Cancers (Basel). 2018;10(6).

427. Hsu JL, Glaser SL. Epstein-barr virus-associated malignancies: epidemiologic patterns and etiologic implications. Crit Rev Oncol Hematol. 2000;34(1):27-53.

428. Public Health Agency of Canada. Brief report: hepatitis B infection in Canada [Available from: http://www.phac-aspc.gc.ca/id-mi/pdf/hepB-eng.pdf].

429. Drolet M, Benard E, Boily MC, Ali H*, et al.* Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. Lancet Infect Dis. 2015;15(5):565-80.

430. Ford AC, Forman D, Hunt RH, Yuan Y*, et al. Helicobacter pylori* eradication therapy to prevent gastric cancer in healthy asymptomatic infected individuals: systematic review and meta-analysis of randomised controlled trials. BMJ. 2014;348:g3174.

431. Volesky KD, El-Zein M, Franco EL, Brenner DR*, et al.* Cancers attributable to infections in Canada. Submitted to Preventive Medicine. 2019.

432. Van de Velde N, Boily MC, Drolet M, Franco EL*, et al.* Population-level impact of the bivalent, quadrivalent, and nonavalent human papillomavirus vaccines: a model-based analysis. J Natl Cancer Inst. 2012;104(22):1712-23.

433. Deeks SL, Tunis MC, Ismail S, National Advisory Committee on I. Summary of the NACI Update on the recommended use of Human Papillomavirus (HPV) vaccine: Nine-valent HPV vaccine two-dose immunization schedule and the use of HPV vaccines in immunocompromised populations. Can Commun Dis Rep. 2017;43(6):138-42.

434. Brisson M, Benard E, Drolet M, Bogaards JA*, et al.* Population-level impact, herd immunity, and elimination after human papillomavirus vaccination: a systematic review and meta-analysis of predictions from transmission-dynamic models. Lancet Public Health. 2016;1(1):e8-e17.

435. Statens Serum Institut. Human papillomavirus-vaccine (HPV) færdigvaccineret - vaccinationstilslutning 2017 [Available from: http://www.ssi.dk/Smitteberedskab/Sygdomsovervaagning.aspx].

436. Hanley SJ, Yoshioka E, Ito Y, Kishi R. HPV vaccination crisis in Japan. Lancet. 2015;385(9987):2571.

437. El-Zein M, Richardson L, Franco EL. Cervical cancer screening of HPV vaccinated populations: Cytology, molecular testing, both or none. J Clin Virol. 2016;76 Suppl 1:S62-S8.

438. Saraiya M, Unger ER, Thompson TD, Lynch CF*, et al.* US assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines. J Natl Cancer Inst. 2015;107(6):djv086.

439. Poirier AE, Ruan R, Walter SD, Franco FL*, et al.* The future burden of cancer in Canada: Long-term cancer incidence projections 2013-2042. Submitted manuscript. 2019.

440. Chandra A, Mosher WD, Copen C, Sionean C. Sexual behavior, sexual attraction, and sexual identity in the United States: data from the 2006-2008 National Survey of Family Growth. Natl Health Stat Report. 2011(36):1-36.

441. Daling JR, Madeleine MM, Johnson LG, Schwartz SM*, et al.* Human papillomavirus, smoking, and sexual practices in the etiology of anal cancer. Cancer. 2004;101(2):270-80.

442. World Health Organization. Global Health Sector Strategy on Viral Hepatitis 2016-2021. Geneva, Switzerland; 2016.

443. Kreimer AR, Struyf F, Del Rosario-Raymundo MR, Hildesheim A*, et al.* Efficacy of fewer than three doses of an HPV-16/18 AS04-adjuvanted vaccine: combined analysis of data from the Costa Rica Vaccine and PATRICIA Trials. Lancet Oncol. 2015;16(7):775-86.

444. Franco E. Guest Editorial: Outlook on the prevention and control of HPV-associated cancers in Canada. In: Canadian Cancer Society's Advisory Committee on Cancer Statistics, editor. Canadian Cancer Statistics 2016. Toronto, ON: Canadian Cancer Society; 2017. p. 99-100.

445. Volesky KD, El-Zein M, Franco EL, Brenner DR*, et al.* Estimates of the future burden of cancer attributable to infections in Canada. Prev Med. 2019;122:118-27.

446. Oh JK, Weiderpass E. Infection and cancer: global distribution and burden of diseases. Ann Glob Health. 2014;80(5):384-92.

447. de Martel C, Maucort-Boulch D, Plummer M, Franceschi S. World-wide relative contribution of hepatitis B and C viruses in hepatocellular carcinoma. Hepatology. 2015;62(4):1190-200.

448. Dalia S, Chavez J, Castillo JJ, Sokol L. Hepatitis B infection increases the risk of non-Hodgkin lymphoma: a meta-analysis of observational studies. Leuk Res. 2013;37(9):1107-15.

Appendix A: List of Publications During Doctoral Studies

1. Volesky KD, Magnan S, Mayrand MH, Isidean SD, El-Zein M, Franco EL, Comète E, Coutlée F. **Clinical performance of the BD Onclarity extended genotyping assay for the management of women positive for human papillomavirus in cervical cancer screening**. Cancer Epidemiology, Biomarkers and Prevention, Published OnlineFirst Feb 7, 2022
2. Mullen CJ, Volesky KD (equally contributing first author), Greenwald Z, El-Zein M, Franco EL. **Is Hodgkin Lymphoma Associated with Hepatitis B and C Viruses? A Systematic Review and Meta-Analysis**. Cancer Epidemiology, Biomarkers, & Prevention. 2021; 30(12):2167-2175
3. Pader J, Ruan Y, Poirier AE, Asakawa K, Lu C, Memon S, Miller A, Walter S, Villeneuve P, King WD, Volesky KD, Smith L, De P, Friedenreich CM, Brenner DR. **Estimates of future cancer mortality attributable to modifiable risk factors in Canada**. Can J Public Health. 2021; 112(6):1069-1082
4. Ruan Y, Poirier AE, Pader J, Asakawa K, Lu C, Memon S, Miller A, Walter S, Villeneuve P, King WD, Volesky KD, Smith L, De P, Friedenreich CM, Brenner DR. **Estimating the future cancer management costs attributable to modifiable risk factors in Canada**. Can J Public Health. 2021; 112(6):1083-1092
5. Franco EL, Shinder GA, Volesky KD, Shapiro SB, MacCosham A. **The noblest among noble public health goals: preventing suicide**. Suicide prevention (special issue). Prev Med. 2021; 152(Pt 1): 106771
6. Franco EL, Shinder GA, Volesky KD, Shapiro SB, MacCosham A. **Lessons from an unparalleled disruption to cancer prevention and control**. From disruption to recovery: The impact of the COVID-19 pandemic on cancer screening (special issue). Prev Med. 2021; 151, 106686
7. Franco EL, Shinder GA, Volesky KD, Shapiro SB, MacCosham A. **A bold but morally necessary and attainable goal**. From science to action to impact: eliminating cervical cancer (special issue). Prev Med. 2021; 144: 106461
8. Malagón T, Volesky KD, Bouten S, Laprise C, El-Zein M, Franco EL. **Cumulative risk of cervical intraepithelial neoplasia for women with normal cytology but positive for human papillomavirus: systematic review and meta-analysis**. Int J of Cancer. 2020; 147(10): 2695-2707
9. Volesky KD, El-Zein M, Franco EL, Brenner DR, Friedenreich CM, Ruan Y, ComPARe Study Team. **Estimates of the future burden of cancer attributable to infections in Canada**. Prev Med. 2019; 122: 118-127
10. Volesky KD, El-Zein M, Franco EL, Brenner DR, Friedenreich CM, Ruan Y, ComPARe Study Team. **Cancers attributable to infections in Canada**. Prev Med. 2019; 122: 109-117
11. Poirier AE, Ruan Y, Volesky KD, King WD, O'Sullivan DE, Gogna P, Walter SD, Villeneuve PJ, Friedenreich CM, Brenner DR, ComPARe Study Team. **The current and future burden of cancer attributable to modifiable risk factors in Canada: Summary of results**. Prev Med. 2019; 122: 140-147
12. Brenner DR, Friedenreich CM, Ruan Y, Poirier AE, Walter SD, King WD, Franco EL, Demers PA, Villeneuve PJ, Grevers X, Nuttall R, Smith L, Volesky KD, O'Sullivan DE, De P, ComPARe Study Team. **The burden of cancer attributable to modifiable risk factors in Canada: Methodological overview**. Prev Med. 2019; 122: 3-8
13. Poirier AE, Ruan Y, Walter SE, Franco EL, Villeneuve PJ, King WD, Volesky KD, O'Sullivan DE, Friedenreich CM, Brenner DR, ComPARe Study Team. **The future burden of cancer in Canada: Long-term cancer incidence projections 2013–2042**. Cancer Epidemiology. 2019; 59: 199-207
14. Brenner DR, Poirier AE, Walter SD, King WD, Franco EL, Demers PA, Villeneuve PJ, Ruan Y, Khandwala F, Grevers X, Nuttall R, Smith L, De P, Volesky KD, O'Sullivan D, Hystad P, Friedenreich CM, ComPARe Study Team. **Estimating the current and future cancer burden in Canada: methodological framework of the Canadian population attributable risk of cancer (ComPARe) study**. BMJ Open. 2018; 8(7): e022378
15. Franco EL, Shinder GA, Tota JE, Volesky KD, Isidean SD. **Journal editors as curators of scholarship: A case study in repairing the scientific record**. Prev Med. 2018; 110: 114-115
16. Volesky KD, Maki A, Scherf C, Watson L, Van Ryswyk K, Fraser B, Weichenthal S, Cassol E, Villeneuve PJ. **The influence of three e-cigarette models on indoor fine and ultrafine particulate matter concentrations under real-world conditions**. Environmental Pollution. 2018; 243(B): 882-889
17. Gagnon MA and Volesky KD. **Merger mania: mergers and acquisitions in the generic drug sector from 1995 to 2016**. Globalization and Health. 2017; 13(1): 1-7
18. Volesky KD and Villeneuve PJ. **Examining screening mammography participation among women aged 40 to 74**. Canadian Family Physician. 2017; 63(6): 300-309

Appendix B: Reprints of Published Manuscripts within Dissertation

Preventive Medicine 122 (2019) 109–117

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Cancers attributable to infections in Canada

Karena D. Volesky[a,b,*], Mariam El-Zein[a], Eduardo L. Franco[a,b], Darren R. Brenner[c], Christine M. Friedenreich[d], Yibing Ruan[d], on behalf of the ComPARe Study Team[1]

[a] Gerald Bronfman Department of Oncology, Division of Cancer Epidemiology, McGill University, Montréal, Québec, Canada
[b] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada
[c] Departments of Oncology and Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[d] Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada

ARTICLE INFO

Keywords:
Cancer
Infection
Population attributable risk
Papillomavirus infections
Helicobacter pylori
Hepatitis viruses
Prevention
Canada

ABSTRACT

Infections are estimated to cause approximately 15% of the world's cancers with large geographic variations. Yet, Canadian estimates for specific cancer-causing infections are not available. To estimate the number of infection-associated cancers diagnosed among Canadian adults in 2015, we calculated population attributable risks (PARs) and the number of attributable cases for seven carcinogenic infections and their 20 associated cancers. A systematic literature search was performed for each infection to obtain data on infection prevalence in the population and the relative risk or odds ratio associated with the cancer it causes. When mechanistic evidence suggested that detection of a given infection within cancer tissue was sufficient to attribute the cancer to the infection, prevalence among cancer cases was used to approximate the PAR. Data from 61 studies formed the basis of our analyses. The estimated number of infection-attributable cancer cases for 2015 was: 3828 for human papillomavirus (HPV), 2052 for *Helicobacter pylori*, 578 for Epstein-Barr virus, 509 for hepatitis B and C viruses (HBV, HCV), 100 for human herpesvirus type 8, and 30 cases for human T-cell lymphotropic virus type 1. These seven infections were responsible for 3.7% of cancers diagnosed among Canadian adults in 2015; 3.5% among men and 4.0% among women. The infections with the highest number of attributable cases are largely preventable or treatable through vaccination (HBV and HPV), antibiotic therapy (*H. pylori*), or a combination of interventions (HCV), thereby representing an important target for reducing the infection-caused cancer burden among Canadians.

1. Introduction

Numerous infectious viruses and bacteria are established risk factors for certain cancers (International Agency for Research on Cancer, 2012). Many carcinogenic infections are strongly associated with specific cancers (e.g., *Helicobacter pylori* (*H. pylori*) and non-cardia gastric cancer, hepatitis B virus (HBV) and hepatitis C virus (HCV) and hepatocellular carcinoma) (Helicobacter and Cancer Collaborative Group, 2001; Cho et al., 2011), while several others are necessary causes for cancer development (e.g., human papillomavirus (HPV) and cervical cancer, human herpesvirus type 8 (HHV-8) and Kaposi sarcoma) (Franco et al., 1999; Mesri et al., 2010).

Globally, almost one-sixth of cancers were attributable to infections with large geographical variations observed (de Martel et al., 2012; Parkin, 2006; Plummer et al., 2016). The proportion of infection-attributable cancers in 2012 varied from a high 31.3% in Sub-Saharan Africa to a low 4.0% in North America (Plummer et al., 2016). Although the latter constitutes a relatively smaller percentage, there is an

* Corresponding author at: Division of Cancer Epidemiology, McGill University, 5100 Maisonneuve Blvd West, Suite 720, Montréal, Québec H4A 3T2, Canada.
E-mail address: karena.volesky@mail.mcgill.ca (K.D. Volesky).

[1] Additional ComPARe Study Team members: Stephen Walter, Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada; Will King, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; Paul Demers, Occupational Cancer Research Centre, Toronto, Ontario, Canada; Paul Villeneuve, Department of Health Sciences, Carleton University, Ottawa, Ontario, Canada; Prithwish De, Cancer Care Ontario, Toronto, Ontario, Canada; Leah Smith, Canadian Cancer Society, Toronto, Ontario, Canada; Abbey Poirier, Department of Cancer Epidemiology and Prevention research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada; Elizabeth Holmes, Canadian Cancer Society, Toronto, Ontario, Canada; Dylan O'Sullivan, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; Zeinab El-Masri, Cancer Care Ontario, Toronto, Ontario, Canada; Robert Nuttall, Canadian Cancer Society, Toronto, Ontario, Canada; Sheila Bouten, Division of Cancer Epidemiology, McGill University, Montréal, Québec, Canada; Tasha Narain, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; Priyanka Gogna, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada.

https://doi.org/10.1016/j.ypmed.2019.03.035

opportunity to lower the Canadian cancer burden with currently available interventions. Specifically, primary preventive interventions include vaccination against HBV and HPV, along with secondary prevention measures such as direct-acting antivirals for chronic HCV infection and antibiotic therapy to treat *H. pylori* infection (De Flora and Bonanni, 2011; Falade-Nwulia et al., 2017; Kohli et al., 2014). The prolonged latency associated with HCV and *H. pylori* provides an opportunity to treat them prior to cancer development (Lingala and Ghany, 2015).

Although, to date, no study has estimated the impact of the different infections on cancer incidence in Canada, a global study reported that 3.9% of incident cancers in Canada were attributable to infections overall in 2012 (Plummer et al., 2016). The global analysis combined infection prevalence for regions comprising many countries; for example, low, medium and high infection incidence areas. Since infection prevalence varies geographically, region-specific data based on more recent evidence from the scientific literature and population-based studies are necessary to obtain accurate estimates of the impact of infections on cancer incidence. Additionally, estimating individually the proportion of cancers attributable to each infection provides essential assessment of the cancer burden due to infections with modifiable prevalence.

Estimates of the impact of each infection on cancer incidence will contribute to the evidence needed to prioritize strategies aimed at reducing the prevalence of certain carcinogenic infections and initiating treatment for others. We estimated, among individuals 18 years and older, the proportion and number of cancers diagnosed in Canada in 2015 that were attributable to infections, by sex and age whenever possible.

2. Methods

The current analysis is part of the ComPARe (Canadian population attributable risk of cancer) Study, which estimates the current and future burden of cancer due to modifiable risk factors in Canada. Here, we estimated the current burden of cancers caused by infections.

2.1. Infections and cancer sites selection

We considered infections classified by the International Agency for Research on Cancer (IARC) as established, Group 1, carcinogens (Table 1). Infections with extremely low prevalence in Canada (*Opisthorchis viverrini*, *Clonorchis sinensis*, and *Schistosoma haematobium*) were excluded. We also did not include human immunodeficiency virus

Table 1
Overview of the carcinogenic infections and associated cancer sites.[a]

Infection	Main transmission route(s)	Main factor(s) for transmission	Carcinogenic mechanism(s)[b] *From Bouvard 2009*	Gold standard for detection	Cancers with *sufficient* evidence[c]	Cancers with *limited* evidence[c]
Hepatitis B virus (HBV), *chronic infection*	Sera and other body fluids	Reusing needles, sexual intercourse	Inflammation Liver cirrhosis Chronic hepatitis	HBsAg	Hepatocellular carcinoma	Cholangiocarcinoma, non-Hodgkin lymphoma
Hepatitis C virus (HCV), *chronic infection*	Sera	Reusing needles	Inflammation Liver cirrhosis Liver fibrosis	HCV RNA	Hepatocellular carcinoma, non-Hodgkin lymphoma	Cholangiocarcinoma
Helicobacter pylori (*H. pylori*)	Oral/fecal	Crowding, contaminated water	Inflammation Oxidative stress Altered cellular turn-over and gene expression Methylation Mutation	Immunoblot	Non-cardia gastric carcinoma, low-grade B-cell MALT gastric lymphoma	None
Epstein-Barr virus[d] (EBV)	Oral/saliva	Pre-chewing food for babies, sharing utensils, kissing	Cell proliferation Inhibition of apoptosis Genomic instability Cell migration	EBER ISH LMP1 IHC for Hodgkin lymphoma (Gulley and Tang, 2008)	Burkitt lymphoma, Hodgkin lymphoma, extranodal natural killer T-cell lymphoma - nasal type, nasopharyngeal carcinoma, immune suppression-related non-Hodgkin lymphoma	Gastric carcinoma, lymphoepithelioma-like carcinoma
Human papillomavirus (HPV), type 16	Skin-to-skin/ mucosal	Sexual contact including oral sex and open mouth kissing	Immortalisation Genomic instability Inhibition of DNA damage response Anti-apoptotic activity	PCR alone or with p16 for anogenital cancers E6 and/or E7 mRNA for head and neck cancers	Cancers of the cervix, anus, penis, vagina, vulva, oropharynx, tonsil, and oral cavity	Laryngeal carcinoma
Human herpesvirus, type 8[e] (HHV-8)	Oral/saliva	Sexual contact including oral sex and open mouth kissing	Cell proliferation Inhibition of apoptosis Genomic instability Cell migration	IFA	Kaposi sarcoma, primary effusion lymphoma	Multicentric Castleman's disease
Human T-cell lymphotropic virus, type 1 (HTLV-1)	Sera and other body fluids, including breast milk	Breast-feeding, sexual intercourse, and reusing needles (Goncalves et al., 2010)	Immortalisation and transformation of T cells	PCR	Adult T-cell leukemia/ lymphoma	None

Abbreviations: HBsAg = Hepatitis B surface antigen, RNA = ribonucleic acid, mRNA = messenger ribonucleic acid, EBER ISH = Epstein-Barr virus encoding region in situ hybridization, LMP1 = latent member protein 1, IHC = immunohistochemistry, PCR = polymerase chain reaction, IFA = immunofluorescent assays, MALT = mucosa-associated lymphoid tissue.

[a] Included infections have been categorized by IARC as Group 1 carcinogens.

[b] Carcinogenic mechanisms were taken from Bouvard 2009 (Bouvard et al., 2009).

[c] Cancer sites were categorized by IARC as having *sufficient* or *limited* evidence.

[d] Epstein-Barr virus is also referred to as human herpesvirus, type 4.

[e] Human herpesvirus, type 8 is also referred to as Kaposi sarcoma virus.

(HIV) because HIV acts indirectly through immunosuppression, thereby amplifying the carcinogenic effects of co-infections such as Epstein-Barr virus (EBV), HCV, and HPV, infections that are already included in our analysis. Table 1 also enumerates the cancers for which there was 'sufficient' evidence for the role of infections in carcinogenesis, as concluded by IARC (International Agency for Research on Cancer, 2012). There was one exception; we estimated the impact of HPV16 on laryngeal cancer incidence because more data have accumulated since the last IARC monograph publication on HPV in support of an etiologic role of HPV in laryngeal cancer (Li et al., 2013; Torrente et al., 2011).

2.2. Population attributable risk calculations

To estimate the proportion of cancer incidence that could have been avoided had the infection been eliminated, we calculated population attributable risks (PARs). The three equations below can estimate PARs for binary exposures (infected or not). The first formula requires the infection prevalence in the general population (Pe) and the relative risk (RR) or odds ratio (OR) associated with the cancer (Levin, 1953). When Pe is not known, the second formula can estimate PARs using prevalence in cases (Pc) in place of Pe (Miettinen, 1974). The third formula is used when the attributable risk in the exposed approaches 1.0 (i.e., RRs are very high), such that the prevalence in cases approximates the PAR.

1. $PAR = \frac{Pe(RR-1)}{1+Pe(RR-1)}$ 2. $PAR = Pc\frac{(RR-1)}{RR}$ 3. $PAR = Pc$

Since we were able to obtain population-based data for HBV, HCV, and *H. pylori* prevalence, the first formula was used for estimating PARs for HBV, HCV, and *H. pylori*. The PARs for the remaining infections, EBV, HPV, HHV-8 and human T-cell lymphotropic virus type 1 (HTLV-1) were estimated with the third formula because they either demonstrate strong relationships with their associated cancers or mechanistic evidence exists for the role of the infection in cancer thus allowing for the PAR to be approximated by the prevalence in cancer cases (International Agency for Research on Cancer, 2012; Plummer et al., 2016; D'Souza et al., 2007).

2.3. Data collection and selection

The data needed to estimate PARs were identified by reviewing IARC monographs (International Agency for Research on Cancer, 2012, 1997, 2007), PAR analyses from other regions (de Martel et al., 2012; Plummer et al., 2016; Antonsson et al., 2015; Parkin, 2011), the Catalan Institute of Oncology HPV Information Centre reports for Canada and the United States (Bruni et al., 2017a; Bruni et al., 2017b), and results of our systematic literature reviews. A systematic literature search was conducted for each infection (details in Supplementary Table 1, S1) to extract data on the infection prevalence and identify meta-analyses on infection-associated cancers. Since the most recent IARC meeting that reviewed each infectious agent considered data published to the end of 2007, we searched for records published in English or French from January 1, 2008 to the search date of June 20, 2017. When data were sparse, we performed more targeted searches in PubMed and contacted experts in their respective fields. Ethics approval was granted for this project by the Health Research Ethics Board of Alberta - Cancer Committee (HREBA.CC-14-0220_REN4), and McGill Univeristy exempted this study from Research Ethics Board review.

Cancers for which the infection is a necessary cause or part of the diagnostic criteria for a given cancer were: cervical cancer, extranodal natural killer T-cell lymphoma - nasal type, Kaposi sarcoma, primary effusion lymphoma, and adult T-cell leukemia/lymphoma, 100% were attributable to their associated infection and therefore inclusion criteria were not required. For all other infections and cancers, the inclusion criteria were: adult population (defined as age 15 and older), North American study population, non-specialized population (e.g. studies performed in exclusively HIV-positive participants were excluded), 10 or more cancer cases, and use of the gold standard method to detect the infection. The inclusion criteria specific to each infection-cancer pair are noted in the tables of included studies (Supplementary Tables 2–13).

When the prevalence in cancer cases approximated the PAR (formula 3), the infection had to be detected in the cancer tumor, such as in a biopsy or surgical specimen. To extrapolate prevalence estimates to recent cancer incidence, rather than incorporating a latency period, the aim was to select studies conducted closer to the timeframe when cancer incidence data were collected. For this reason, studies had to be published in 1995 or later. Specifically, the prevalence of any HPV in the oropharynx has increased over time in the USA; pre-1990 HPV prevalence was 20.9% and from 2000 to 2013 it rose to 65.4% (Stein et al., 2014), further emphasizing the importance of utilizing more recent studies.

The prevalence of HBV and *H. pylori* were derived from North American population-based serosurveys, and HCV prevalence was extracted from a study that modeled chronic HCV prevalence in the Canadian population (Trubnikov et al., 2014). Due to limited data on the measures of association for *H. pylori* associated cancers, a posteriori decision was made to consider studies conducted among European populations and studies that used the detection method that preceded the current gold standard method (we corrected to the new standard).

The chosen detection method for assessing the presence of infection was crucial to the PAR estimation. Selecting studies that utilized the gold standard detection method was prioritized over other factors such as having a Canadian population or sex and age-specific results leading to sparser data.

2.4. Estimating infection prevalence in the Canadian population

Below is a brief description of how we adjusted population-based data to obtain sex- and age-specific estimates of HBV, HCV, and *H. pylori* prevalence for the Canadian population. The prevalence estimates and further details are provided in supplementary Tables S2–S5.

2.5. Hepatitis B virus

The Canadian Health Measures Survey (CHMS) was the first population-based survey to provide estimates of HBV and HCV prevalence for the Canadian population (Rotermann et al., 2013). Data from two cycles of the CHMS (Statistics Canada, n.d.), collected from 2007 to 2009 and 2009 to 2011, were combined for the analysis. The combined participation rate for those providing direct health measures after sample strategy adjustments was 52.8% for the two cycles (Rotermann et al., 2013). Sera from CHMS participants aged 14–79 testing positive for hepatitis B core antigen (anti-HBc) were then tested for hepatitis B surface antigen (HBsAg). Chronic HBV infection is defined as the presence of HBsAg six months after a positive HBV test (National Notifiable Diseases Surveillance System, 2012). Given the cross-sectional design of the CHMS, we assumed that HBsAg positivity at one time point represented chronic HBV infection. Privacy restrictions limited HBsAg results to either sex or broad age groups (14–49 and 50–79), yet sex and age effect HBV prevalence. To obtain Canadian age-specific prevalence estimates, we used the HBsAg 10-year age-group prevalence from two merged cycles of the weighted National Health and Nutrition Examination Survey (NHANES) (Centers for Disease Control and Prevention, 2009, 2011) to partition the CHMS estimates by 10-year age groups. The first two cycles of the CHMS were collected from 2007 to 2011, resulting in a six-year latency. This time period does not correspond to the prolonged latency for hepatocellular carcinoma (El-Serag, 2012), yet it is still plausible as the CHMS measured prevalent not incident HBV infection.

2.6. Hepatitis C virus

The CHMS is a household-based survey of non-institutionalized populations (Statistics Canada, 2010). Thus, groups with higher HCV prevalence, namely intravenous drug users, were underrepresented by excluding those who were homeless or in prison. Moreover, although a diagnosis of either HBV or HCV in Canada are reported to national public health agencies (Public Health Agency of Canada, 2018), many of these infections remain undiagnosed and therefore uncaptured in this data source. We thus obtained the modeled chronic HCV prevalence by birth cohort from Trubnikov, Yan and Archibald who accounted for high-risk groups and undiagnosed infections in their analyses (Trubnikov et al., 2014). To obtain chronic HCV prevalence by sex, we partitioned the estimates using the sex distribution of HCV prevalence from another study that modeled HCV prevalence in Canada in 2007 (Remis, 2010). Since the latency period between initial HCV infection and hepatocellular carcinoma is 25–30 years (Lingala and Ghany, 2015), we used the midpoint of a 15-year latency in our estimates.

We did not estimate a PAR for HBV and HCV coinfection and hepatocellular carcinoma because data on coinfection prevalence were not available. To estimate the combined impact of HBV and HCV on hepatocellular carcinoma, we combined their PARs with the following equation: 1 – (1-HBV PAR) * (1-HCV PAR) (Miettinen, 1974).

2.7. Helicobacter pylori

Few studies have assessed *H. pylori* prevalence in Canadian populations. Although most of these studies were conducted with specialized populations (Cheung et al., 2014; Sethi et al., 2013), one study included 1306 residents aged 50–80 in Canada's most populous province, Ontario (Naja et al., 2007). As population-based data covering a broad age range were required, we opted to utilize other data. *H. pylori* serostatus was assessed in one NHANES cycle collected from 1999 to 2000 (Centers for Disease Control and Prevention, 2001) which resulted in a 15–16 year latency period. The weighted NHANES data were re-weighted by sex, five-year age groups, and race/ethnicity (Black, Latin American, White, and Other) to better reflect the composition of the Canadian population in 2001 (the closest year for which Canadian census ethnicity data were available). Missing *H. pylori* results, accounting for 5.0–6.6% of the reweighted data, were assumed to be missing completely at random and excluded. Additionally, half of the 1–2% 'equivocal' results, which were the results of IgG levels between the cut-offs for positive and negative results, were re-assigned as positive or negative. NHANES used enzyme-linked immunosorbent assay (ELISA) to detect *H. pylori*. ELISA has a sensitivity of 95.6% and specificity of 92.6% (Monteiro et al., 2001). We corrected our reweighted prevalence data according to these reported diagnostic accuracy measures (Franco, 1992).

Since immunoblot is more sensitive than ELISA for the detection of *H. pylori* in gastric cancer cases (Gonzalez et al., 2012; Mitchell et al., 2008), we also corrected the association measures from matched case-control studies that used ELISA by deriving a formula used to adjust the OR, (Franco, 1992) and calculating sensitivity and specificity parameters. The latter were derived by pooling the sensitivity and specificity from three studies (Gonzalez et al., 2012; Mitchell et al., 2008; Peleteiro et al., 2010), that directly compared ELISA and immunoblot in the same patients.

2.8. Estimating infection prevalence in cancer cases

The PARs for EBV- and HPV- associated cancers were approximated by pooling studies that provided data on the prevalence of the infectious agent as detected in cancer tissues. For anogenital cancers, we considered an infection with at least one high-risk type (HPVs 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73, and 97) to indicate that the cancer was due to HPV. Head and neck cancers were considered attributable to HPV if genotype 16 was found via the detection of E6 and/or E7 oncoproteins which indicates viral activity and replication.

2.9. Cancer incidence

To determine the number of cases attributable to a given infection, the calculated PAR is multiplied by the number of incident cases. Incident cancer data were obtained from the Canadian Cancer Registry for 2015, which was the most recent year available. When data were requested for rare or subsite cancers, they were aggregated to maintain privacy; for example, cancer incidence counts were combined into two age groups (ages < 50, and ≥ 50), instead of five-year age groups. To preserve the granularity in the incidence data, we estimated the proportion of liver cancer estimated to be hepatocellular carcinoma. A study using SEER (Surveillance, Epidemiology, and End Results) data reported that there were 55,344 primary liver cancers diagnosed from 1978 to 2007, of which 44,080 were hepatocellular carcinoma (Altekruse et al., 2011). We applied the ensuing proportion of 0.797 (44,080/55,344) to liver cancer incidence to get the estimated number of hepatocellular carcinoma cases.

For the province of Quebec, the most recent year for which cancer incidence data were available was 2010. Quebec's 2015 cancer incidence was estimated in one of two ways. For cancers with fewer cases (< 500 in Canada in 2015), the last five years of available incidence data for Quebec, 2006–2010, were averaged and applied to Quebec's 2015 population. For other cancers, Quebec's 2015 incidence was imputed by fitting a Poisson regression on Canada's 2008–2015 incidence.

2.10. Statistical analysis

To obtain the prevalence of a given infection in its associated cancer, individual studies were pooled with a random effects model; a fixed effect model was adopted if the index of consistency (I^2) was $<$ 25% and if the test for heterogeneity was not statistically significant ($p > 0.05$). To pool the proportions and measures of association, we used the commands *metaprop* (Nyaga et al., 2014) and *metan* (Harris et al., 2008), respectively. To calculate 95% confidence intervals (CIs) for the pooled proportions, the exact method was used with the command "cimethod (exact)". When studies were excluded by the software because of inadmissible 95% CIs (e.g. proportions of 1.0 can yield CIs over 1.0), the Freeman-Tukey double arcsine transformation was enabled to calculate admissible 95% CIs bounded by 0.0–1.0 (stata command is: "ftt"). All meta-analyses were conducted in Stata v14 (StataCorp., College Station, TX, USA). R was used to calculate PARs via formula 1 (R Foundation for Statistical Computing [Internet], 2017). For infections where the PAR was approximated by the prevalence of the infection in cancer cases, no additional calculations were necessary after pooling the prevalence. The CIs for PARs calculated via formula 1 were calculated as previously described (Brenner et al., 2018; Brenner et al., 2019).

3. Results

A summary of the overall methods and findings for HBV, HCV, *H. pylori* is presented in Table 2, and for infections where the prevalence in cases approximated the PAR in Table 3. Specific results and tabulations on the characteristics of included studies as well as forest plots, are provided under the respective infection and cancer sites (Supplementary Tables 6–13 and Figs. 1–8).

Table 2 shows that the prevalence of chronic HBV infection in the Canadian population was $< 1.0\%$ across all age and sex groups whereas chronic HCV prevalence ranged from 0.1 to 1.9%. The prevalence of *H. pylori* was notably lower among those younger than 50 years (12.8% for men and 9.8% for women) compared to those aged 50 years and over (27.9% for men and 29.6% for women). Between 1.6 and 15.3% of

Table 2
Infections for which the attributable risk was estimated using the prevalence in the population and measures of association.

Infection cancer (ICD-03 code)	Method of infection measurement	Source of prevalence data	Range of prevalence estimates by sex	Data used to estimate measure of association	Odds ratio (95% CI)
Helicobacter pylori					
Stomach, non-cardia (C16.1–16.9)	Serology with ELISA or immunoblot detection	NHANES (1999–2000) data reweighted by Canada's sex, age, and race/ethnicity distribution.	Men: 12.8% (aged < 50) to 27.9% (aged ≥50)	Pooled unadjusted ORs from matched case-control studies with fixed effects: 3 corrected studies that used ELISA and 3 studies that used immunoblot	9.4 (6.5–13.4)
Stomach, MALT lymphoma (9699)	Serology with ELISA detection	Estimates were corrected for sensitivity and specificity.	Women: 9.8% (aged < 50) to 29.6% (aged ≥50)	One study of 33 cases matched to 134 controls (Parsonnet et al., 1994)	6.3 (2.0–19.9)
Hepatitis B virus					
Hepatocellular carcinoma (C22.0, 817)	Serology with HBsAg detection	CHMS HBsAg data (2007–2011) partitioned with NHANES HBsAg 10-year age group distribution (2007–2010)	Men: 0.1% (aged 70–79) to 0.9% (aged 30–39) Women: 0.1% (aged 70–79) to 0.7% (aged 30–39)	Meta-analysis with pooled estimate from 3 case-control studies conducted in the USA and 1 cohort study from Australia (Cho et al., 2011)	20.3 (11.3–36.5)
Hepatitis C virus					
Hepatocellular carcinoma (C22.0, 817)	Estimates from modeling studies (Trubnikov et al., 2014; Remis, 2010)	Chronic HCV prevalence modeled for the Canadian population by five-year birth cohorts, partitioned with the sex distribution from another modeling study	Men: 0.2% (aged 16–20) to 1.9% (aged 46–50) Women: 0.1% (aged 16–20) to 1.2% (aged 46–50)	Pooled from seven studies from the USA and Australia (Cho et al., 2011)	23.8 (16.9–33.5)
Non-Hodgkin lymphoma (9591)				Adjusted OR calculated from SEER Medicare data with 33,940 cases matched to controls on sex, age, and year of diagnosis (Anderson et al., 2008)	1.35 (1.06–1.73)

Abbreviations: CI = confidence interval, MALT = mucosa-associated lymphoid tissue, NHANES = National Health and Nutrition Examination Survey, CHMS = Canadian Health Measures Survey, HBsAg = Hepatitis B surface antigen, SEER = Surveillance, Epidemiology, and End Results (United States).

Table 3
Methods used for the infections where population attributable risks were estimated using the prevalence of infection in cancer cases.

Infection cancer (ICD-03 code)	Method of infection measurement	Source of prevalence estimates[d]	Cases used to estimate PAR, n	Sex/age group	PAR (prevalence of infection in cancer cases)	
					Estimate (%)	95% CI
Epstein-Barr virus						
Burkitt lymphoma (9687)	EBER ISH	1 study	30	< 50 years old	40.0	22.7–59.4
			21	≥50 years old	28.6	11.3–52.2
ENKTL, nasal type (9719)			–	All	100.0	–
Hodgkin lymphoma (C81)	EBER ISH and/or LMP1 IHC	4 studies	560	Men	43.0	28.4–57.7
			583	Women	26.6	12.1–41.1
Nasopharynx (C11)	EBER ISH	2 studies	172	All	69.4	61.9–76.9
Human papillomavirus, high-risk types,[a] anogenital tract cancers						
Anus (C21)	PCR detection with genotyping of at least HPV 16 and 18	5 studies	154	Men	87.6	76.4–95.8
			250	Women	94.6	89.3–98.3
Cervix (C53)		Necessary cause	–	Women	100.0	–
Penis (C60)		6 studies	311	Men	39.3	21.8–56.9
Vagina (C52)		2 studies	85	Women	72.2	62.7–81.7
Vulva (C51)		2 studies	43	< 50 years old	76.8	64.2–89.4
		3 studies	201	≥50 years old	43.2	13.9–72.5
Human papillomavirus, type 16, head and neck cancers						
Oropharynx[b] (C01.9, C02.8, C02.4, C05.1, C05.2, C14.2, C09, C10)	PCR with E6 and/or E7 for HPV16	16 studies	1396	All	60.2	51.8–68.5
Oral cavity[c] (C00.4–0.5, C00.9, C02.0–C02.9, C03, C04, C05.0, C05.8, C05.9, C06, C14.8)		9 studies	733	All	8.2	3.6–14.2
Larynx (C32)		5 studies	194	All	12.7	3.7–25.4
Human herpesvirus, type 8						
Kaposi sarcoma (9140)	IFA	Necessary cause	–	All	100.0	–
Primary effusion lymphoma (9678)	IFA	Part of diagnostic criteria	–	All	100.0	–
Human T-cell lymphotropic virus, type 1						
Adult T-cell leukemia and lymphoma (9827)	PCR	Necessary cause	–	All	100.0	–

Abbreviations: EBER ISH = EBV-encoded RNA in situ hybridization, PCR = polymerase chain reaction, LMP1 = latent member protein 1, IHC = immunohistochemistry, CI = confidence interval, PAR = population attributable risk, ENKTL = extranodal natural killer T-cell lymphoma, IFA = immunofluorescent assays.

[a] High-risk HPV types include types classified by the International Agency for Research on Cancer as Group 1 (16, 18, 31, 33, 35, 39, 45, 51, 56, 58 and 59), Group 2A (68) and Group 2B (34, 53, 66, 70 and 73) carcinogens. HPV types 52 and 97 were also considered high-risk types.

[b] Oropharynx subsites: base of the tongue (C01.9), overlapping lesion of tongue (C02.8), lingual tonsil (C02.4), soft palate (C05.1), uvula (C05.2), Waldeyer ring (C14.2), tonsil (C09), oropharynx (C10).

[c] Oral cavity subsites: mucosa of lip (C00.4–0.5) and lip NOS (C00.9), other and unspecified parts of tongue (C02.0–C02.9), gum (C03), floor of mouth (C04), palate - hard, overlapping lesion, NOS (C05.0, C05.8, C05.9), other and unspecified parts of mouth (C06) and overlapping lesion of lip, oral cavity and pharynx (C14.8).

[d] Included studies can be found in the supplement under their respective infection and cancers.

hepatocellular carcinomas were attributable to chronic HBV infection (Supplement, Table S2). Chronic HCV had higher attributable percentages than HBV, ranging from 2.5 to 30.0% (Supplement, Table S4). However, the percent of non-Hodgkin lymphoma attributable to HCV was negligible (< 0.7%) for each age and sex group.

As shown in Table 3, the proportion of cancer attributable to high-risk HPV types in anogenital cancers was lowest for penile cancer (39.3%) and highest for cervical cancer (100.0%). The presence of HPV16 in head and neck cancers was 60.2% for the oropharynx, 12.7% for the larynx and 8.2% for the oral cavity.

Table 4 demonstrates that HPV infections were the causative agent for most infection-associated cancers (3828, 95% CI: 3190–4425), followed by *H. pylori* (2052, CI: 1473–2395), and EBV (578, CI: 286–604). More than half (54.0%) of the infection-caused cancers diagnosed in 2015 were related to HPV, then *H. pylori* (28.9%), EBV (8.1%), HBV/HCV (7.2%), HHV-8 (1.4%), and finally HTLV-1 (0.4%) (data not shown). The cancers with the highest number of attributable cases were: non-cardia stomach (n = 1730), cervix (n = 1375), oropharynx (n = 1083), anus (n = 589), and hepatocellular carcinoma (n = 480) (Table 4). A total of 7097 cancers were attributable to infections, representing an estimated 3.7% of the cancers diagnosed among those ≥18 years old in 2015. The proportion of incident cancers attributable to infections was higher among women (4.0%) than men (3.5%).

4. Discussion

The proportion of attributable cancers in Canada in 2015 ranged from a low of 0.4% for HCV in non-Hodgkin lymphoma to a high of 100.0% for HPV in cervical cancer. Cervical cancer was one of five cancers where all cases are attributable to an infection. With few exceptions (HCV in non-Hodgkin lymphoma, and HPV in the oral cavity and larynx), all the calculated PARs exceeded 25.0%, thereby demonstrating the important role that infections play in certain malignancies.

We found that the burden of infection-caused cancers was higher among women (4.0%) than men (3.5%), largely because of HPV's role in cervical and other anogenital cancers. Estimates for the United Kingdom also demonstrated a higher attributable proportion among women than men (3.7% versus 2.5%, respectively) in 2011 (Parkin, 2011) and a similar finding was found in Australia where 2.4% of cancers diagnosed among men in 2010 were attributed to infections and 3.7% among women (Antonsson et al., 2015). In contrast, an analysis for the USA found that 3.3% among both men and women were attributable to infections in 2014 (Islami et al., 2018).

As PAR estimates assume causality between the exposure and outcome, we included only established carcinogens and cancers where the evidence for the role of the infection was deemed 'sufficient' by the IARC (except for HPV16 in laryngeal cancer). Yet, there is increasing

Table 4
Summary of the number of cases and proportion of cancers attributable to infections in Canada in 2015

Infection, cancer(s)	Total			Men			Women		
	Obs cases[a]	AC[b]	% Attributable[c]	Obs cases	AC	% Attributable	Obs cases	AC	% Attributable
Hepatitis B and C virus									
Hepatocellular carcinoma	1750	480	27.4	1345	400	29.7	405	80	19.8
Hepatitis C virus									
Non-Hodgkin lymphoma	8290	29	0.4	4620	19	0.4	3670	10	0.3
Helicobacter pylori									
Stomach, MALT lymphoma	560	322	57.5	265	151	57.0	295	171	58.0
Stomach, non-cardia	2515	1730	68.8	1445	993	68.7	1070	737	68.9
Epstein-Barr virus									
Burkitt lymphoma	85	30	35.3	65	23	35.4	20	7	35.0
ENKTL – nasal type	25	25	100.0	15	15	100.0	10	10	100.0
Hodgkin lymphoma	940	336	35.8	525	226	43.0	415	110	26.6
Nasopharynx	270	187	69.4	195	135	69.4	75	52	69.4
Human papillomavirus, high-risk types									
Anus	640	589	92.0	225	197	87.6	415	392	94.5
Cervix	1375	1375	100.0	…	…	…	1375	1375	100.0
Penis	205	81	39.3	205	81	39.3	…	…	…
Vagina	180	130	72.2	…	…	…	180	130	72.2
Vulva	635	301	47.4	…	…	…	635	301	47.4
Human papillomavirus, type 16									
Oropharynx[h]	1800	1083	60.2	1380	830	60.2	420	253	60.2
Oral cavity	1560	127	8.2	940	77	8.2	620	51	8.2
Larynx	1115	142	12.7	925	118	12.7	190	24	12.7
Human herpesvirus, type 8									
Kaposi sarcoma	90	90	100.0	70	70	100.0	20	20	100.0
Primary effusion lymphoma	10	10	100.0	10	10	100.0	…	…	100.0
Human T-cell lymphotropic virus, type 1									
Adult T-cell leukemia and lymphoma	30	30	100.0	15	15	100.0	15	15	100.0
All associated cancers[d]	22,075	7097	32.2	12,245	3360	27.4	98,30	3738	38.0
All cancers[g]	189,530	7097	3.7	96,070	3360	3.5	93,460	3738	4.0

Abbreviations: Obs = observed, AC = attributable cases, MALT = mucosa-associated lymphoid tissue, ENKTL = extranodal natural killer T-cell lymphoma.

[a] Cancer incidence data for the year 2015 from the Canadian Cancer Registry. Quebec's cancer incidence was estimated. Hepatocellular carcinoma incidence was estimated by applying the proportion 0.797 to liver cancer incidence.

[b] Number of cancer cases at individual cancer sites that can be attributed to infection.

[c] Proportion attributable was calculated by dividing the number of cases attributable to infection by the number of the associated cancer cases. It differs from PAR which for some cancers varied by age and/or sex.

[d] All associated cancers includes all cancers known to be associated with infections listed in the table.

[g] All cancers includes all incident cancer cases in Canada among those 18 and older in 2015.

[h] Includes the base of the tongue and tonsils.

evidence that other infection cancer associations including EBV in gastric carcinoma, HBV in non-Hodgkin lymphoma and HCV in cholangiocarcinoma, among others, may also cause cancer. If these associations were included, the impact of infections on cancer incidence would have been higher than what we reported here.

4.1. Hepatitis B and C viruses, and H. pylori

The combined impact of the hepatitis viruses resulted in 27.4% of hepatocellular carcinoma incidence being attributable to HBV/HCV. Since HBV can be avoided with vaccination that began in Canada in the early 1980s, and because HCV can be prevented through a variety of behavioral interventions and treated with direct-acting antivirals, the future burden of hepatocellular carcinoma has the potential to decrease by reducing the prevalence of these viruses.

Globally, *H. pylori* was responsible for 89% of non-cardia gastric cancers (Plummer et al., 2015). We calculated that 68.8% of incident non-cardia gastric cancers in Canada were due to this infection. We estimated PARs based on elimination of the infection. This information is helpful for understanding the impact of infections on cancer incidence; however, in practice, elimination may not be entirely feasible. For example, *H. pylori* can be treated with quadruple antibiotic therapy, but challenges in the scalability of screening for the infection and concerns over antibiotic resistance limit the prospect of eliminating the infection at the population level (Bourke et al., 2005; Hunt et al., 2004; Fallone et al., 2016).

4.2. EBV, HHV-8 and HTLV-1

Although EBV is the infection with the highest prevalence with > 90% of adults infected (de-The et al., 1975), it was responsible for only 8.1% of the infection-caused cancers in Canada in 2015. In a similar vein, some infections with PARs of 100% were responsible for a small number of cancers (e.g. HHV-8 and HTLV-1) because of the rarity of cancers they cause.

4.3. Human papillomavirus

We found that 54% of the infection-associated cancers were due to HPV. This percentage is higher than the reported 29.5% global contribution of HPV to infection-associated cancers (Plummer et al., 2016). In particular our estimates for HPV16's role in head and neck cancers were higher than global estimates. Meta-analyses have reported higher HPV prevalence in oropharyngeal cancers in North American populations compared to other continents (Ndiaye et al., 2014; Mehanna et al., 2013). Our estimate of 60.2% with E6/E7 detection, albeit numerically similar to that of Ndiaye et al. (60.4%) (Ndiaye et al., 2014), is actually higher than the latter because it represents detection of HPV16, whereas the 60.4% estimate in that study is for all HPV types combined. The oropharynx had the third highest number of attributable cases. Since 1997, oropharyngeal and oral cancer incidence has increased in Canada, especially among men, this is in part due to HPV's role in head and neck cancers (Canadian Cancer Society's Advisory Committee on

Cancer Statistics, 2015). The Canadian Cancer Society estimated that in 2012, cervical and oropharyngeal cancers each accounted for 35% of the HPV-associated cancer burden. We too, found that approximately one-third of the HPV associated cancer burden was due to cervical (35.9%) and oropharyngeal (28.3%) cancers. Since we examined the contribution of HPV16, any of the three available HPV vaccines provide coverage against this HPV type. Although a smaller proportion of oral cavity and laryngeal cancers are attributable to HPV16 (8.2% and 12.7%, respectively), they added 269 cases to the infection-associated cancer burden. School-based HPV immunization programs began in Canada in 2007. More recently, these programs have been extended to boys (Shapiro et al., 2017). We found that one-third (34.0%) of HPV associated cancers were diagnosed among men, which further emphasizes the importance of vaccinating boys.

4.4. Limitations

The main limitation of our study was the lack of Canadian-specific infection data and the subsequent reliance on data collected in the United States and for *H. pylori* data collected from European populations. We have assumed that the exposure prevalence and strength of the relationship between the infection and cancer as observed in American and European populations were comparable to what would have been observed in Canada. For example, we reweighted the age, sex and race/ethnicity distribution from a population-based survey of *H. pylori* prevalence in the United States (NHANES) to match that of the Canadian population in the closest available year. Reweighting assumed that differences in the prevalence of *H. pylori* between the two countries were due to age, sex, and race/ethnicity – but these variables do not likely fully account for the potential differences between Canada and the United States. For some infection cancer site pairs, irrespective of including data collected outside of Canada and performing more targeted literature searches, the data remained sparse. This situation was particularly true for: *H. pylori* and gastric mucosa-associated lymphoid tissue lymphoma, EBV and Burkitt lymphoma, and HPV and vaginal cancer. This result was anticipated since the cancer sites with sparser evidence were also the rarer cancers.

Focusing exclusively on Canada allowed us to obtain much of the rare and subsite cancer incidence data we required for accurate estimates of the number of attributable cases. However, we estimated rather than directly obtained hepatocellular carcinoma and Quebec's cancer incidence. For cancer sites with fewer than 500 cases in Canada in 2015, the five-year incidence rates were averaged but this averaging relies on assumptions that the average of the last five years of available cancer incidence for Quebec (2006–2010) is representative of the 2015 cancer incidence, and that the trend has remained stable.

Since we used existing data, our findings inherited the methodologic flaws of included studies and population-based surveys. This concern was at least partially mitigated by including only those studies that met stringent inclusion criteria aimed at enhancing the validity of our estimates. We attempted to correct for measurement error; however, some error may remain. Additionally, our correction for error in assessing the association between *H. pylori* and non-cardia gastric cancer did not account for confounders. Although the included studies were matched case-control studies, unmatched confounders have not been adjusted for.

By not conducting a separate analysis for HIV, we potentially underestimated the impact of infections on cancer incidence. The proportion of cancer attributable to EBV has the potential to increase since non-Hodgkin lymphomas among HIV positive populations were not included in this analysis.

5. Conclusion

We estimated that 3.7% of cancers diagnosed among Canadians aged 18 and older in 2015 were attributable to seven carcinogenic infections. This percentage translated into 7097 cancers, where ~6400 could potentially be prevented with currently available vaccines or treatments. HPV was responsible for more cancers than other infections, comprising more than half of the infection-associated cancer burden. The presence of three vaccines that confer 95% efficacy against the HPV types responsible for cancer incidence is encouraging (Kash et al., 2015). Although Canada has a lower infection-associated cancer burden relative to many other countries (Plummer et al., 2016), infection-associated cancers continue to impact cancer incidence and increasing vaccine hesitancy has the potential to limit the progress that could be made in reducing the HPV and HBV associated cancer burden.

Acknowledgements

Karena D. Volesky is supported by a doctoral award from the Fonds de recherche du Quebec - Santé and the Fondation des étoiles. Darren Brenner holds a Canadian Cancer Society Prevention Capacity Development Award (#703917). Christine Friedenreich was supported by an Alberta Innovates Health Senior Scholar Award and by the Alberta Cancer Foundation Weekend to End Women's Cancers Breast Cancer Chair. We thank Shemay Lee for assisting us with data management.

Competing interests

None.

Disclosure

E.L.F. has served as occasional consultant to companies involved with HPV diagnostics and vaccination (Merck, GSK, Roche, and BD). His institution has received grants from Merck and Roche to supplement investigator-initiated studies that he leads at McGill University.

E.L.F. is Editor-in-Chief at Preventive Medicine and K.D.V. is an Assistant Editor at Preventive Medicine. The process of soliciting the special issue, sending out manuscripts for review, the peer-review process and editorial decision making was conducted entirely outside of the Preventive Medicine online system (for which E.L.F. and K.D.V. have access to through their regular Preventive Medicine duties).

Funding sources

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

Role of funding source

The sponsor had no role in the study design, data collection, analysis or interpretation of the data, writing of the manuscript or in the decision to submit this manuscript for publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ypmed.2019.03.035.

References

Altekruse, S.F., Devesa, S.S., Dickie, L.A., McGlynn, K.A., Kleiner, D.E., 2011. Histological classification of liver and intrahepatic bile duct cancers in SEER registries. J. Registry Manag. 38 (4), 201–205.

Anderson, L.A., Pfeiffer, R., Warren, J.L., Landgren, O., Gadalla, S., Berndt, S.I., et al., 2008. Hematopoietic malignancies associated with viral and alcoholic hepatitis. Cancer Epidemiol. Biomark. Prev. 17 (11), 3069–3075.

Antonsson, A., Wilson, L.F., Kendall, B.J., Bain, C.J., Whiteman, D.C., Neale, R.E., 2015. Cancers in Australia in 2010 attributable to infectious agents. Aust. N. Z. J. Public Health 39 (5), 446–451.

Bourke, B., Ceponis, P., Chiba, N., Czinn, S., Ferraro, R., Fischbach, L., et al., 2005. Canadian Helicobacter Study Group Consensus Conference: update on the approach

to Helicobacter pylori infection in children and adolescents–an evidence-based evaluation. Can. J. Gastroenterol. 19 (7), 399–408.
Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., et al., 2009. A review of human carcinogens—part B: biological agents. Lancet Oncol. 10 (4), 321–322.
Brenner, D.R., Poirier, A.E., Walter, S.D., King, W.D., Franco, E.L., Demers, P.A., et al., 2018. Estimating the current and future cancer burden in Canada: methodological framework of the Canadian population attributable risk of cancer (ComPARe) study. BMJ Open 8 (7).
Brenner, D.R., Friedenreich, C.M., Ruan, Y., Poirier, A.E., Walter, S.D., King, W.D., et al., 2019. The Burden of Cancer Attributable to Modifiable Risk Factors in Canada: Methodological Overview. 122. pp. 3–8.
Bruni, L.B.-R.L., Albero, G., Serrano, B., Mena, M., Gómez, D., Muñoz, J., Bosch, F.X., de Sanjosé, S., 2017a. Human Papillomavirus and Related Diseases in USA. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre).
Bruni, L.B.-R.L., Albero, G., Serrano, B., Mena, M., Gómez, D., Muñoz, J., Bosch, F.X., de Sanjosé, S., 2017b. Human Papillomavirus and Related Diseases in Canada. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre).
Canadian Cancer Society's Advisory Committee on Cancer Statistics, 2015. Canadian Cancer Statistics. (Toronto, ON).
Centers for Disease Control and Prevention, 2001. In: Department of Health and Human Services (Ed.), NHANES 1999–2000 Laboratory Data.
Centers for Disease Control and Prevention, 2009. In: Department of Health and Human Services (Ed.), NHANES 2007–2008 Laboratory Data.
Centers for Disease Control and Prevention, 2011. In: Department of Health and Human Services (Ed.), NHANES 2009–2010 Laboratory Data.
Cheung, J., Goodman, K.J., Girgis, S., Bailey, R., Morse, J., Fedorak, R.N., et al., 2014. Disease manifestations of Helicobacter pylori infection in Arctic Canada: using epidemiology to address community concerns. BMJ Open 4 (1), e003689.
Cho, L.Y., Yang, J.J., Ko, K.P., Park, B., Shin, A., Lim, M.K., et al., 2011. Coinfection of hepatitis B and C viruses and risk of hepatocellular carcinoma: systematic review and meta-analysis. Int. J. Cancer 128 (1), 176–184.
De Flora, S., Bonanni, P., 2011. The prevention of infection-associated cancers. Carcinogenesis. 32 (6), 787–795.
de Martel, C., Ferlay, J., Franceschi, S., Vignat, J., Bray, F., Forman, D., et al., 2012. Global burden of cancer attributable to infections in 2008: a review and synthetic analysis. Lancet 13, 607–615.
de-The, G., Day, N.E., Geser, A., Lavoue, M.F., Ho, J.H., Simons, M.J., et al., 1975. Sero-epidemiology of the Epstein-Barr virus: preliminary analysis of an international study - a review. IARC Sci. Publ. 11 (Pt 2), 3–16.
D'Souza, G., Kreimer, A.R., Viscidi, R., Pawlita, M., Fakhry, C., Koch, W.M., et al., 2007. Case-control study of human papillomavirus and oropharyngeal cancer. N. Engl. J. Med. 356 (19), 1944–1956.
El-Serag, H.B., 2012. Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology. 142 (6), 1264–73 e1.
Falade-Nwulia, O., Suarez-Cuervo, C., Nelson, D.R., Fried, M.W., Segal, J.B., Sulkowski, M.S., 2017. Oral direct-acting agent therapy for hepatitis C virus infection: a systematic review. Ann. Intern. Med. 166 (9), 637–648.
Fallone, C.A., Chiba, N., van Zanten, S.V., Fischbach, L., Gisbert, J.P., Hunt, R.H., et al., 2016. The Toronto Consensus for the treatment of Helicobacter pylori infection in adults. Gastroenterology. 151 (1), 51–69 e14.
Franco, E.L., 1992. Measurement errors in epidemiological studies of human papillaomvirus and cervical cancer. In: Muñoz, N.B., F.X., Meheus, A. (Eds.), The Epidemiology of Human Papillomavirus and Cervical Cancer. International Agency for Reseach on Cancer, Lyon, pp. 119.
Franco, E.L., Rohan, T.E., Villa, L.L., 1999. Epidemiologic evidence and human papillomavirus infection as a necessary cause of cervical cancer. J. Natl. Cancer Inst. 91 (6), 506–511.
Goncalves, D.U., Proietti, F.A., Ribas, J.G., Araujo, M.G., Pinheiro, S.R., Guedes, A.C., et al., 2010. Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases. Clin. Microbiol. Rev. 23 (3), 577–589.
Gonzalez, C.A., Megraud, F., Buissonniere, A., Lujan Barroso, L., Agudo, A., Duell, E.J., et al., 2012. Helicobacter pylori infection assessed by ELISA and by immunoblot and noncardia gastric cancer risk in a prospective study: the Eurgast-EPIC project. Ann. Oncol. 23 (5), 1320–1324.
Gulley, M.L., Tang, W., 2008. Laboratory assays for Epstein-Barr virus-related disease. J. Mol. Diagn. 10 (4), 279–292.
Harris, R.B.M., Deeks, J., Harbord, R., Altman, D., Sterne, J., 2008. metan: fixed- and random-effects meta-analysis. Stata J. 8 (1), 3–28.
Helicobacter and Cancer Collaborative Group, 2001. Gastric cancer and Helicobacter pylori: a combined analysis of 12 case control studies nested within prospective cohorts. Gut. 49 (3), 347–353.
Hunt, R., Fallone, C., Veldhuyzan van Zanten, S., Sherman, P., Smaill, F., Flook, N., et al., 2004. Canadian Helicobacter Study Group Consensus Conference: update on the management of Helicobacter pylori—an evidence-based evaluation of six topics relevant to clinical outcomes in patients evaluated for H pylori infection. Can. J. Gastroenterol. 18 (9), 547–554.
International Agency for Research on Cancer, 1997. Epstein-Barr Virus and Kaposi's Sarcoma Herpesvirus/Human Herpesvirus 8. World Health Organization.
International Agency for Research on Cancer, 2007. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Human Papillomavirus.
International Agency for Research on Cancer, 2012. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Biological Agents. Lyon, France.
Islami, F., Goding Sauer, A., Miller, K.D., Siegel, R.L., Fedewa, S.A., Jacobs, E.J., et al., 2018. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. CA Cancer J. Clin. 68 (1), 31–54.
Kash, N., Lee, M., Kollipara, R., Downing, C., Guidry, J., Tyring, S., 2015. Safety andefficacy data on vaccines and immunization to human papillomavirus. J. Clin.Med. 4, 614–633.
Kohli, A., Shaffer, A., Sherman, A., Kottilil, S., 2014. Treatment of hepatitis C: a systematic review. JAMA. 312 (6), 631–640.
Levin, M.L., 1953. The occurrence of lung cancer in man. Acta Unio Int. Contra Cancrum 9 (3), 531–541.
Li, X., Gao, L., Li, H., Gao, J., Yang, Y., Zhou, F., et al., 2013. Human papillomavirus infection and laryngeal cancer risk: a systematic review and meta-analysis. J. Infect. Dis. 207 (3), 479–488.
Lingala, S., Ghany, M.G., 2015. Natural history of hepatitis C. Gastroenterol. Clin. N. Am. 44 (4), 717–734.
Mehanna, H., Beech, T., Nicholson, T., El-Hariry, I., McConkey, C., Paleri, V., et al., 2013. Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer—systematic review and meta-analysis of trends by time and region. Head Neck 35 (5), 747–755.
Mesri, E.A., Cesarman, E., Boshoff, C., 2010. Kaposi's sarcoma and its associated herpesvirus. Nat. Rev. Cancer 10 (10), 707–719.
Miettinen, O.S., 1974. Proportion of disease caused or prevented by a given exposure, trait or intervention. Am. J. Epidemiol. 99 (5), 325–332.
Mitchell, H., English, D.R., Elliott, F., Gengos, M., Barrett, J.H., Giles, G.G., et al., 2008. Immunoblotting using multiple antigens is essential to demonstrate the true risk of Helicobacter pylori infection for gastric cancer. Aliment. Pharmacol. Ther. 28 (7), 903–910.
Monteiro, L., de Mascarel, A., Sarrasqueta, A.M., Bergey, B., Barberis, C., Talby, P., et al., 2001. Diagnosis of Helicobacter pylori infection: noninvasive methods compared to invasive methods and evaluation of two new tests. Am. J. Gastroenterol. 96 (2), 353–358.
Naja, F., Kreiger, N., Sullivan, T., 2007. Helicobacter pylori infection in Ontario: prevalence and risk factors. Can. J. Gastroenterol. 21 (8), 501–506.
National Notifiable Diseases Surveillance System, 2012. Hepatitis B, Chronic, 2012 Case Definition: Centers for Disease Control and Prevention. Available from: https://wwwn.cdc.gov/nndss/conditions/hepatitis-b-chronic/case-definition/2012/.
Ndiaye, C., Mena, M., Alemany, L., Arbyn, M., Castellsague, X., Laporte, L., et al., 2014. HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: a systematic review and meta-analysis. Lancet Oncol. 15 (12), 1319–1331.
Nyaga, V.N., Arbyn, M., Aerts, M., 2014. Metaprop: a Stata command to perform meta-analysis of binomial data. Arch Public Health 72 (1), 39.
Parkin, D., 2006. The global health burden of infection-associated cancers in the year 2002. Int. J. Cancer 118, 3030–3044.
Parkin, D.M., 2011. Cancers attributable to infection in the UK in 2010. Br. J. Cancer 105 (Suppl. 2), S49–S56.
Parsonnet, J., Hansen, S., Rodriguez, L., Gelb, A.B., Warnke, R.A., Jellum, E., et al., 1994. Helicobacter pylori infection and gastric lymphoma. N. Engl. J. Med. 330 (18), 1267–1271.
Peleteiro, B., Lunet, N., Barros, R., La Vecchia, C., Barros, H., 2010. Factors contributing to the underestimation of Helicobacter pylori-associated gastric cancer risk in a high-prevalence population. Cancer Causes Control 21 (8), 1257–1264.
Plummer, M., Franceschi, S., Vignat, J., Forman, D., de Martel, C., 2015. Global burden of gastric cancer attributable to Helicobacter pylori. Int. J. Cancer 136 (2), 487–490.
Plummer, M., de Martel, C., Vigant, J., Ferlay, J., Bray, F., Franceschi, A., 2016. Global burden of cancer attributable to infections in 2012: a synthetic analysis. Lancet 4 (9) (e609–e16).
Public Health Agency of Canada, 2018. List of Nationally Notifiable Diseases. Government of Canada Available from: http://diseases.canada.ca/notifiable/diseases-list.
R: a language and environment for statistical computingR Foundation for Statistical Computing [Internet] Available from: https://www.R-project.org/.
Remis, R., 2010. Modelling the Incidence of Prevalence of Hepatitis C Infection and Its Sequelae in Canada, 2007. Public Health Agency of Canada.
Rotermann, M., Langlois, K., Andonov, A., M T., 2013. Seroprevalence of hepatitis B and C virus infections: results from the 2007 to 2009 and 2009 to 2011 Canadian Health Measures Survey. Statistics Canada. Health Rep. 24 (11), 3–13 82-003-X.
Sethi, A., Chaudhuri, M., Kelly, L., Hopman, W., 2013. Prevalence of Helicobacter pylori in a First Nations population in northwestern Ontario. Can. Fam. Physician 59 (4), e182–e187.
Shapiro, G.K., Guichon, J., Kelaher, M., 2017. Canadian school-based HPV vaccine programs and policy considerations. Vaccine. 35 (42), 5700–5707.
Statistics Canada, 2010. Canadian Health Measures Survey (CHMS) Data User Guide: Cycle 1.
Statistics Canada Canadian Health Measures Survey (CHMS). Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5071.
Stein, A.P., Saha, S., Yu, M., Kimple, R.J., Lambert, P.F., 2014. Prevalence of human papillomavirus in oropharyngeal squamous cell carcinoma in the United States across time. Chem. Res. Toxicol. 27 (4), 462–469.
Torrente, M.C., Rodrigo, J.P., Haigentz Jr., M., Dikkers, F.G., Rinaldo, A., Takes, R.P., et al., 2011. Human papillomavirus infections in laryngeal cancer. Head Neck 33 (4), 581–586.
Trubnikov, M., Yan, P., Archibald, C., 2014. Estimated prevalence of hepatitis C virus infection in Canada, 2011. Can. Commun. Dis. Rep. 40 (19), 429–436.

Preventive Medicine 122 (2019) 118–127

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Estimates of the future burden of cancer attributable to infections in Canada

Karena D. Volesky[a,b,*], Mariam El-Zein[a], Eduardo L. Franco[a,b], Darren R. Brenner[c], Christine M. Friedenreich[d], Yibing Ruan[d], on behalf of the ComPARe Study Team[1]

[a] Gerald Bronfman Department of Oncology, Division of Cancer Epidemiology, McGill University, Montréal, Québec, Canada
[b] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada
[c] Departments of Oncology and Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[d] Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada

ARTICLE INFO

Keywords:
Cancer
Infection
Potential impact fraction
Papillomavirus infections
Helicobacter pylori
Hepatitis viruses
Prevention
Canada

ABSTRACT

More than 7000 incident cancers diagnosed in Canada in 2015 were attributable to infections. The future infection-associated cancer burden can be lowered by reducing the prevalence of major cancer-causing infections; hepatitis B virus (HBV), hepatitis C virus (HCV), *Helicobacter pylori* (*H. pylori*) and human papillomavirus (HPV). We modeled the future impact of (1) 10%, 25%, and 50% relative reductions in the prevalence of HBV, HCV and *H. pylori* and (2) different school-based HPV vaccination coverage levels (lower, current, higher) on Canadian cancer incidence by the year 2042. We modeled counterfactual reductions in HBV, HCV and *H. pylori* prevalence in 2018, assuming a latency period of 15-years, to estimate the impact on cancer incidence starting in 2033. The number of HPV-attributable cancers among vaccinated cohorts was a function of pre-2018 vaccine coverage levels and the 2018 counterfactuals. A 50% counterfactual reduction in the prevalence of HBV, HCV and *H. pylori* could prevent an estimated 10,585 cancers from 2018 to 2042; a 25% reduction could prevent 5293 cancers and a 10% reduction could prevent 2117 cancers. Assuming continuity of current estimated country-wide HPV vaccine coverage, 3977 anogenital and 1073 head and neck cancers could be prevented from 2018 to 2042, whereas vaccine coverage of 80% in girls and boys could prevent an additional 311 cancers. Almost 16,000 cancers could be prevented in Canada from 2018 to 2042 with a 50% relative reduction in HBV, HCV and *H. pylori* prevalence and 80% HPV vaccine coverage of girls and boys.

1. Introduction

Globally, an estimated 14.0% of cancers diagnosed in 2012 were attributable to four infectious agents; hepatitis B virus (HBV), hepatitis C virus (HCV), *Helicobacter pylori* (*H. pylori*) and human papillomavirus (HPV) (Plummer et al., 2016). Several strategies have been adopted to reduce the prevalence of cancer-causing infections and their associated cancer or pre-cancer incidence in Canada and abroad. Canadian provinces/territories introduced publicly-funded, school-based immunization programs for HBV from 1992 to 1998 and for HPV from 2007 to 2010 (Government of Canada, 2017; Shapiro et al., 2017). Due to HBV's long latency, reductions in cancer incidence have not yet been realized. However, the annual number of reported HBV infections in Canada has decreased from 10.8 per 100,000 persons in 1990 to 1.7 per 100,000 persons in 2008 (Public Health Agency of Canada, 2011). A meta-analysis of 20 ecologic population-based studies conducted in high-income countries reported a 68% decrease in the prevalence of HPV types 16 and 18 at a vaccination coverage among girls of 50% or higher (Drolet et al., 2015). A meta-analysis of randomized controlled trials reported that eradication of *H. pylori* in asymptomatic populations

* Corresponding author at: Division of Cancer Epidemiology, McGill University, 5100 Maisonneuve Blvd West, Suite 720, Montreal, Québec H4A 3T2, Canada.
E-mail address: karena.volesky@mail.mcgill.ca (K.D. Volesky).

[1] Additional ComPARe Study Team members: Stephen Walter, Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada; Will King, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; Paul Demers, Occupational Cancer Research Centre, Toronto, Ontario, Canada; Paul Villeneuve, Department of Health Sciences, Carleton University, Ottawa, Ontario, Canada; Prithwish De, Cancer Care Ontario, Toronto, Ontario, Canada; Leah Smith, Canadian Cancer Society, Toronto, Ontario, Canada; Abbey Poirier, Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada; Dylan O'Sullivan, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; Zeinab El-Masri, Cancer Care Ontario, Toronto, Ontario, Canada; Robert Nuttall, Canadian Cancer Society, Toronto, Ontario, Canada; Sheila Bouten, Gerald Bronfman Department of Oncology, Division of Cancer Epidemiology, McGill University, Montréal, Québec, Canada; Tasha Narain, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; Priyanka Gogna, Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada.

https://doi.org/10.1016/j.ypmed.2019.04.006

reduced the relative risk of gastric cancer by 34% (Ford et al., 2014).

Despite infections' impact on global cancer incidence, the level of public awareness of a causal role for infections in the development of cancer is low. Yet, the public plays a key role by vaccinating their children against HBV and HPV, not reusing needles and complying with antibiotic treatment for *H. pylori* infection. The range of primary prevention strategies aimed at reducing the acquisition of infections (HBV, HCV and HPV) and secondary prevention strategies for treating existing infections (HCV, *H. pylori*) provides an opportunity to lower the infection-associated cancer burden.

We estimated that > 7000 cases of cancers, representing 3.7% of all cancers diagnosed among Canadians aged 18 and older in 2015 were attributable to seven carcinogenic infections (Volesky et al., 2019). The vast majority (90.0%) of these infection-attributable cancers were due to HBV, HCV, *H. pylori* and HPV. We found that, with 3828 attributable cases, more cancers were attributed to HPV than any other infection. The infection with the next highest number of attributable cases was *H. pylori* with 2052 cases, followed by Epstein-Barr virus with 578 cases, hepatitis B and C viruses with 509 cases, human herpesvirus type 8 (i.e. Kaposi sarcoma virus) with 100 cases and finally human T-cell lymphotropic virus type 1 with 30 attributable cases in 2015.

The considerable potential to prevent carcinogenic infections highlights the importance of quantifying the impact of a variety of prevention scenarios, referred to as counterfactuals, for prioritizing strategies aimed at reducing the number of infection-associated cancers. To our knowledge, besides the impact of HPV on cancer incidence (Van de Velde et al., 2012), no study has estimated the impact of reductions in the prevalence of infections on the future Canadian cancer incidence. We estimated the future burden of infection-associated cancers by the year 2042 by modeling the impact of: 1) relative reductions in HBV, HCV and *H. pylori* infection prevalence and 2) lower, current, and higher levels of school-based HPV vaccination coverage.

2. Methods

This analysis is part of the Canadian population attributable risk of cancer (ComPARe) Study, which aimed to estimate the current and future burden of cancer attributable to modifiable risk factors in Canada (Brenner et al., 2018). Here, we estimated the future burden of cancers caused by four major infectious agents (HBV, HCV, *H. pylori* and HPV). The future burden and the potential for prevention of infection-associated cancers are reported as: the number of cancers projected and prevented in 2042 and the cumulative number of cancers prevented from 2018 to 2042 based on different counterfactuals.

We calculated potential impact fractions (PIFs) to estimate the proportion of HBV, HCV and *H. pylori*-associated incident cancers that could be avoided by 2042 under various counterfactual scenarios, using the following equation (Morgenstern and Bursic, 1982):

$$PIF = \frac{(P–P^*)(RR–1)}{P(RR–1) + 1}$$

where P is the pre-counterfactual infection prevalence, P* is the post-counterfactual infection prevalence, and RR is the relative risk or odds ratio (OR) for the association between the infection and cancer. The annual prevented cases were estimated as:

$$PC_i = I_i \times PIF$$

where I_i is the projected cancer incidence in year *i*.

For HPV, we approximated the proportion of cancers attributable to HPV by using prevalence of HPV in cancer cases and therefore did not calculate PIFs. Knowing the proportion of specific cancers attributable to HPV enabled us to estimate the number of avoidable HPV-related cancer cases. When estimating the future number of preventable cancers among vaccinated cohorts, the proportion attributable to HPV was multiplied by the relevant cancer incidence, after accounting for vaccine efficacy, protection (e.g. the proportion of HPV types contributing to cancer incidence that are covered by the vaccines), and coverage.

2.1. Current infection prevalence

We have reported on the prevalence of chronic HBV and HCV, and *H. pylori* for the Canadian population elsewhere (Volesky et al., 2019). Briefly, chronic HBV prevalence (measured by hepatitis B surface antigen (HBsAg)), was assessed using data from two merged cycles (2007–2009 and 2009–2011) of the Canadian Health Measures Survey (CHMS) (Statistics Canada, n.d.; Statistics Canada, 2010). Since we were only able to obtain sex-specific prevalence estimates from the CHMS, HBsAg prevalence from two merged cycles of the United States' National Health and Nutrition Examination Survey (NHANES) were used to partition the HBsAg sex prevalence estimates from the CHMS by 10-year age groups (Centers for Disease Control and Prevention, 2009, 2011). To estimate chronic HCV prevalence, we partitioned the five-year birth cohort estimates from a modeling study (Trubnikov et al., 2014) according to the sex distribution reported in a study that modeled acute and chronic HCV prevalence in the Canadian population (Remis, 2010). Since we required that prevalence estimates originate from population-based data covering a range of ages, the few studies assessing *H. pylori* sero status in Canadian populations did not meet this criterion (Cheung et al., 2014; Naja et al., 2007; Sethi et al., 2013). Hence, to estimate the prevalence of *H. pylori*, we reweighted NHANES data collected from 1999 to 2000 (Centers for Disease Control and Prevention, 2001) to reflect the Canadian age, sex, and race/ethnic composition (categories available were: Black, Latin American, White, and Other). To produce summary prevalence estimates, we calculated population-weighted prevalence estimates by sex thereby aggregating prevalence across age-groups (Table 1).

Rather than estimating HPV prevalence among the Canadian population, we estimated HPV prevalence among cancer cases. Since mechanistic evidence indicates that the detection of high-risk HPV types within cancer tissue is sufficient to attribute that cancer to HPV, the population attributable risk (PAR) was approximated by the prevalence in cases. The prevalence of HPV infection was calculated by pooling, using a random effects model, the proportion of cancer cases harboring high-risk HPV types (for anogenital cancers) or HPV16 (for head and neck cancers) within the cancer tumor tissue. We restricted our analyses to studies that applied “gold standard” HPV detection techniques: polymerase chain reaction (PCR) for anogenital cancers and detection of E6 and/or E7 oncoproteins via PCR for head and neck cancers (Bishop et al., 2012; Rietbergen et al., 2013).

Table 1 summarizes the prevalence of these infections in the population (for HBV, HCV and *H. pylori*) or cancer cases (for HPV), the RRs or ORs and attributable percentages used in our analyses.

2.2. Future infection prevalence

We assumed a constant prevalence of HBV (from 2007 to 2011) and *H. pylori* (from 1999 to 2000) up till 2027. We projected the future chronic HCV prevalence based on prevalence at three time points (1999, 2004, and 2009). Chronic HCV prevalence at these time points was estimated by weighting the available five-year birth cohort data (Trubnikov et al., 2014) by Canada's population to obtain the weighted average prevalence for Canadians aged 15 to 70. To project the future chronic HCV prevalence, an exponential regression was fit between the estimated prevalence and the three time points.

For the baseline HPV prevalence projections, we also assumed no change in prevalence given the lack of evidence in support of an increasing or decreasing trend in prevalence within cases. Although the prevalence of HPV within oropharyngeal cancer has increased over time (Stein et al., 2014), mostly due to a decrease in cigarette smoking, we assumed that this trend would not continue post-2018.

Table 1
Cancer types and proportions attributable to carcinogenic infections with modifiable prevalence in Canada[a].

Infection Cancer sites (ICD-03 codes)	Prevalence of the infection in the population, %[b]	Odds ratio or relative risk	Attributable, %[c]	
			Men	Women
Hepatitis B virus (HBV), chronic infection				
Hepatocellular carcinoma (C22, 817)	0.54 (men) 0.36 (women)	20.3	9.5	6.5
Hepatitis C virus (HCV), chronic infection				
Hepatocellular carcinoma (C22, 817)	1999: 1.09 (men) and 0.73 (women)	23.4	16.0	11.3
Non-Hodgkin lymphoma (9591)	2004: 1.05 (men) and 0.70 (women) 2009: 0.99 (men) and 0.66 (women)	1.4	0.3	0.2
Helicobacter pylori (*H. pylori*)				
Gastric non-cardia (C16.1–16.9)	18.0 (men)	9.4	60.0	59.0
Gastric MALT lymphoma (9699)	17.2 (women)	6.3	48.8	47.7

Infection Cancer sites (ICD-03 codes)	Prevalence of the infection in cancer cases, %	Odds ratio or relative risk	Attributable, %[c]	
			Men	Women
Human papillomavirus (HPV), high-risk types[d]		Not applicable as the prevalence in cancer cases approximates the proportion attributable to the infection.		
Cervix (C53)	100.0			
Anus (C21)	87.6 (men) 94.5 (women)			
Penis (C60)	39.4			
Vagina (C52)	72.2			
Vulva (C51)	76.8 (aged 18–49 years) 43.2 (aged ≥50 years)			
Human papillomavirus (HPV), type 16				
Oropharynx (C10, C01, C09)[e]	60.2			
Oral cavity (C02, C03, C04, C06)[f]	8.2			
Larynx (C32)	12.7			

Abbreviations: MALT = mucosa-associated lymphoid tissue.
[a] Detailed description of the prevalence and odds ratio/relative risk estimates can be found in Volesky et al. (2019).
[b] The prevalence of the infection in the population was calculated by weighting the age-group specific prevalence estimates by the Canadian population for each sex.
[c] The attributable percent by sex was calculated by dividing the number of attributable cases by the number of cases, and hence it does not reflect the proportion attributable by specific age groups.
[d] High-risk HPV types include types classified by the International Agency for Research on Cancer as Group 1 (16, 18, 31, 33, 35, 39, 45, 51, 56, 58 and 59), Group 2A (68) and Group 2B (34, 53, 66, 70 and 73) carcinogens. HPV52 and 97 were also considered high-risk types.
[e] Oropharynx subsites: oropharynx (C10), base of the tongue (C01), and tonsil (C09).
[f] Oral cavity subsites: gum and other mouth (C03, C06), floor of mouth (C04), other and unspecified parts of tongue (C02).

2.3. Counterfactual scenarios

We projected the impact of four counterfactual scenarios: no change in the prevalence of HBV and *H. pylori* and a continuing trend for HCV, as well as 10%, 25% and 50% reductions in infection prevalence. These reductions were selected to respectively represent plausible minor, moderate and major prevalence reductions. The counterfactuals were "implemented" in the year 2018 with a 15-year latency to observe an impact on cancer incidence starting in 2033.

There is no treatment for an HPV infection; it can be cleared by the immune system rather than by an intervention (Bosch et al., 2013). We purposely ignored the impact of cervical cancer screening in achieving further cervical cancer incidence reduction and thus selected counterfactuals based on HPV vaccination coverage in girls only, and in girls and boys. Canada's National Advisory Committee on Immunization recommended HPV vaccination for girls in 2007 and for boys in 2012 (Deeks et al., 2017). We considered several plausible counterfactuals for school-based HPV vaccination starting in 2018: 1) maintenance of current coverage among girls, 2) decrease in coverage among girls (40%, 50%, and 60% coverage) and 3) increase in coverage among girls only to 80% and 4) an 80% coverage of school-aged girls and boys (which is sufficient for the elimination of HPV16 (Brisson et al., 2016)), both as direct effects (i.e. those who were vaccinated are protected and no one else) and as herd effects (i.e. vaccine protection extends beyond those directly immunized). A decrease in coverage was considered for two reasons. First, a 50% coverage, although lower than the national average, is the current coverage level in certain regions of Canada (Shapiro et al., 2017). In addition, some countries such as Denmark and Japan have experienced substantial decreases in their coverage levels due to unconfirmed reports of adverse events (Statens Serum Institut, 2017; Hanley et al., 2015). For example, in Sapporo, Japan, the reported three-dose HPV vaccination completion rate ranged from 68.4 to 74.0% and two years later it dropped to 0.6% (Hanley et al., 2015). For comparison, we also present the expected cancer incidence that could have occurred had the HPV vaccine never been administered at any point in time.

2.4. Latency period

HBV, HCV, and *H. pylori* are associated with prolonged latencies that can span decades before cancer diagnosis (El-Serag, 2012; Lingala and Ghany, 2015). For these infections, we assumed a 15-year interval between the time of prevalence reduction and its impact on the incidence of associated cancers; a shorter latency was an appropriate approach given that the data captured prevalent (recent and persistent) rather than incident infections. For HCV, the available data did not allow for direct estimation of the prevalence among individuals 70 years of age or older, so we allowed for a longer latency (between 15 and 20 years) in this age-group. For HPV-associated cancers, we did not account for a latency period because we utilized a cohort approach in which five-year age group cohorts (i.e. 20–24, 25–29, etc.) were followed through time to 2042.

Table 2
Sex-specific projected number of cancer cases and potential impact fractions for chronic hepatitis B and C viruses that could be prevented in 2042 under different counterfactuals.

Infection and associated cancer	Sex	Future burden measures	No change	Cancer burden by reductions in infection prevalence		
				10%	25%	50%
Hepatitis B virus	Men	Projected in 2042	2640	2615	2578	2516
Hepatocellular carcinoma		PIF, %	–	0.9	2.4	4.7
		Prevented in 2042	0	25	62	125
		Prevented 2018–2042	0	110	275	551
	Women	Projected in 2042	718	713	706	695
		PIF, %	–	0.6	1.6	3.2
		Prevented in 2042	0	5	12	23
		Prevented 2018–2042	0	20	50	100
	Both	Projected in 2042	3358	3329	3284	3210
		PIF, %	–	0.9	2.2	4.4
		Prevented in 2042	0	30	74	148
		Prevented 2018–2042	0	130	326	651
Hepatitis C virus	Men	Projected in 2042	2640	2598	2535	2429
Hepatocellular carcinoma		PIF, %	–	1.6	4.0	8.0
		Prevented in 2042	0	42	106	212
		Prevented 2018–2042	0	190	476	952
	Women	Projected in 2042	718	710	697	677
		PIF, %	–	1.1	2.8	5.6
		Prevented in 2042	0	8	20	41
		Prevented 2018–2042	0	36	90	179
	Both	Projected in 2042	3358	3308	3232	3106
		PIF, %	–	1.5	3.8	7.5
		Prevented in 2042	0	50	126	252
		Prevented 2018–2042	0	226	565	1131
Hepatitis C virus	Men	Projected in 2042	5850	5849	5846	5842
Non-Hodgkin lymphoma		PIF, %	–	0.0	0.1	0.1
		Prevented in 2042	0	2	4	9
		Prevented 2018–2042	0	8	19	39
	Women	Projected in 2042	4750	4749	4748	4745
		PIF, %	–	< 0.1	< 0.1	0.1
		Prevented in 2042	0	1	2	5
		Prevented 2018–2042	0	4	10	21
	Both	Projected in 2042	10,600	10,598	10,594	10,587
		PIF, %	–	0	0.1	0.1
		Prevented in 2042	0	3.0	7.0	13.0
		Prevented 2018–2042	0	12	30	59
Hepatitis C virus	Men	Projected in 2042	8491	8447	8381	8271
Total		PIF, %	–	0.5	1.3	2.6
		Prevented in 2042	0	44	110	220
		Prevented 2018–2042	0	198	495	990
	Women	Projected in 2042	5468	5459	5445	5423
		PIF, %	–	0.2	0.4	0.8
		Prevented in 2042	0	9	23	45
		Prevented 2018–2042	0	40	100	200
	Both	Projected in 2042	13,959	13,905	13,826	13,693
		PIF, %	–	0.4	1.0	1.9
		Prevented in 2042	0	53	133	265
		Prevented 2018–2042	0	238	595	1190

Abbreviations: PIF = potential impact fraction.

2.5. Human papillomavirus model parameters

2.5.1. Start date of vaccine coverage

School-based immunization of girls in grades 4 to 7 was introduced in Canadian provinces from 2007 to 2010. Specifically, Ontario (Canada's most populous province) started vaccinating grade 7 girls in 2007, whereas Quebec began vaccinating grade 4 girls in 2008 and British Columbia started vaccinating grade 6 girls in 2008 (Shapiro et al., 2017). We selected the year 2008 as the single start date for country-wide vaccination of girls, which corresponds to the median year when vaccination began. School-based catch-up HPV vaccination programs were extended to boys, first in Prince Edward Island (province with the smallest population) in 2013 and, to a few other jurisdictions (province/territory) in the following years. As we are not accounting for catch-up vaccination here, we did not consider the impact of catch-up vaccination targeted at boys prior to 2018.

2.5.2. Current vaccine coverage

To estimate current Canada-wide vaccine coverage across jurisdictions, we calculated a weighted proportion based on average vaccine completion rates (receiving the last dose of a two or three dose schedule) for the available school years within each jurisdiction (Shapiro et al., 2017). The weights were represented by the proportion of girls aged 10–14 in a particular jurisdiction relative to their Canadian counterparts for the year 2014. The weights were based on the 2014 population levels because vaccine completion rates were reported for school years ranging from 2011/12 to 2015/16. Country-wide coverage was estimated because we lacked provincial level cancer incidence data for some HPV-associated cancer sites (e.g. vagina, vulva, base of tongue and tonsil), and provincial cancer incidence could only be projected to 2038 due to smaller sample size hindering stable projections past 2038. We calculated the school-based vaccination completion rate for Canada using a weighted mean based on the size of each province's proportion of the female Canadian population aged 10–14 years as weights. The

Table 3
Sex-specific projected number of cancer cases and potential impact fractions for *Helicobacter pylori* that could be prevented in 2042 under different counterfactuals.

Cancer	Sex	Future burden measures	No change	Cancer burden by reductions in infection prevalence		
				10%	25%	50%
Gastric MALT lymphoma	Men	Projected in 2042	1389	1321	1219	1050
		PIF, %	–	4.9	12.2	24.4
		Prevented in 2042	0	68	170	339
		Prevented 2018–2042	0	272	679	1358
	Women	Projected in 2042	1014	966	893	772
		PIF, %	–	4.8	11.9	23.8
		Prevented in 2042	0	48	121	242
		Prevented 2018–2042	0	198	496	992
	Both	Projected in 2042	2403	2287	2112	1822
		PIF, %	–	4.8	12.1	24.2
		Prevented in 2042	0	116	290	581
		Prevented 2018–2042	0	470	1175	2351
Gastric non-cardia cancer	Men	Projected in 2042	2823	2654	2399	1976
		PIF, %	–	6.0	15.0	30.0
		Prevented in 2042	0	170	424	848
		Prevented 2018–2042	0	717	1792	3585
	Women	Projected in 2042	2274	2140	1939	1604
		PIF, %	–	5.9	14.7	29.5
		Prevented in 2042	0	134	335	670
		Prevented 2018–2042	0	562	1404	2809
	Both	Projected in 2042	5097	4794	4338	3579
		PIF, %	–	6.0	14.9	29.8
		Prevented in 2042	0	304	759	1518
		Prevented 2018–2042	0	1279	3197	6393
Total	Men	Projected in 2042	4212	3975	3619	3025
		PIF, %	–	5.6	14.1	28.2
		Prevented in 2042	0	237	593	1187
		Prevented 2018–2042	0	989	2472	4943
	Women	Projected in 2042	3288	3105	2832	2376
		PIF, %	–	5.5	13.9	27.7
		Prevented in 2042	0	182	456	912
		Prevented 2018–2042	0	760	1900	3801
	Total	Projected in 2042	7500	7080	6450	5401
		PIF, %	–	5.6	14.0	28.0
		Prevented in 2042	0	420	1049	2099
		Prevented 2018–2042	0	1749	4372	8744

Abbreviations: PIF = potential impact fraction, MALT = mucosa-associated lymphoid tissue.

resulting estimate, 72.4% among girls, was imputed to 2008, which was approximately the median year when school-based programs were introduced.

2.5.3. Vaccine efficacy and protection

Efficacy against high-grade cervical, vaginal, and vulvar disease/cancer based on per-protocol analyses of HPV vaccination trials was reported to range from 95% to 100% in HPV-naïve populations (FUTURE II Study Group, 2007; Huh et al., 2017). To be conservative, we used 95% efficacy in our calculations. Currently, three HPV vaccines are available; the most cancer causing HPV types covered by these vaccines are 16 and 18 (in bi/quadrivalent and nonavalent), and the nonavalent also protects against types 31, 33, 45, 52 and 58. Since the nonavalent vaccine has been in use in all Canadian jurisdictions as of 2018, we modeled its use starting in 2018 for the other counterfactuals. For cohorts vaccinated prior to 2018, we assumed that the quadrivalent vaccine was administered.

With respect to cervical cancer, we utilized protection levels of 70.8% for the bi/quadrivalent and 89.5% for the nonavalent vaccines since these proportions represent the estimated relative contribution of HPV types covered by the respective vaccines (de Martel et al., 2017). Since we had previously estimated the proportion of anogenital cancers due to high-risk HPV types (Table 1), we calculated the proportion of high-risk HPV types included in the vaccines to determine their associated level of protection. For this estimation, we relied on data from a study that reported HPV type distribution in anogenital cancer specimens obtained from population-based registries in the United States (Saraiya et al., 2015). Specifically, to determine the level of vaccine protection, we estimated the proportion of the identified high-risk types covered by the bi/quadrivalent and nonavalent vaccines. We estimated that among high-risk HPV positive cancers, protection of the bi/quadrivalent vaccines ranged from 66.0% (vaginal cancer) to 87.1% (anal cancer), and nonavalent protection ranged from 94.3% (penile cancer) to 97.7% (anal cancer).

2.5.4. Herd immunity

The HPV vaccine confers different levels of herd immunity among non-vaccinated girls and boys. We extracted and interpolated herd effects from a modeling study that meta-analyzed transmission-dynamic models from high-income countries (Brisson et al., 2016). Brisson et al. calculated that 40% vaccine coverage of girls would produce 53% protection among women and 36% among men whereas for 80% coverage of girls, 93% protection among women and 83% among men would be observed (Brisson et al., 2016). For 50%, 60% and 72.4% vaccine coverage levels, we assumed that herd effects would increase by 10% increments. For example, a 50% coverage of girls would produce an estimated effect of 63% (10% higher than the 53% herd effect reported for 40% coverage of girls) and 46% coverage of boys (10% higher than the 36% herd effect for boys when 40% of girls are vaccinated). For current coverage of 72.4%, we increased the herd effect by an additional 2.4% to match the increase in coverage from 60% to 72.4%.

2.5.5. Estimating preventable cases

To determine the proportion of future cancer incidence that could be prevented under the different HPV vaccine coverage counterfactuals,

Table 4
Projected number of anogenital cancers and the number that could be prevented according to variations in school-based HPV vaccine coverage in Canada[a,b,c].

Cancer site, sex	Future burden measures	Lower (%)							Current (%)		Higher (%)		
	Effect:	Direct [d]	Direct	Herd	Direct	Herd	Direct	Herd	Direct	Herd	Direct	Herd	Herd
	Girls:	0.0	40.0	53.0	50.0	63.0	60.0	73.0	72.4	85.4	80.0	93.0	100.0
	Boys:	0.0	0.0	36.0	0.0	46.0	0.0	56.0	0.0	68.4	0.0	83.0	100.0
Cervix	Projected in 2042	1939	1723	1684	1693	1654	1663	1624	1626	1587	1603	1564	1543
	Prevented in 2042	0	216	255	246	285	276	315	313	352	336	375	396
	Prevented 2018–2042	0	2813	2980	2941	3108	3070	3236	3228	3395	3326	3492	3583
Anus, men[e]	Projected in 2042	345	345	345	345	345	345	344	345	344	345	344	343
	Prevented in 2042	0	0	0	0	0	0	0	0	1	0	1	1
	Prevented 2018–2042	0	0	8	0	9	0	10	0	11	0	13	25
Anus, women	Projected in 2042	775	758	757	757	756	756	755	755	754	755	754	753
	Prevented in 2042	0	17	18	18	19	19	20	20	21	20	21	22
	Prevented 2018–2042	0	130	131	131	132	132	133	133	135	134	136	136
Anus, both	Projected in 2042	1120	1103	1102	1102	1101	1101	1100	1100	1099	1100	1098	1097
	Prevented in 2042	0	17	18	18	19	19	20	20	21	20	22	23
	Prevented 2018–2042	0	130	139	131	141	132	143	133	146	134	148	161
Penis	Projected in 2042	260	260	258	260	258	260	258	260	258	260	258	258
	Prevented in 2042	0	0	2	0	2	0	2	0	2	0	2	2
	Prevented 2018–2042	0	0	13	0	14	0	14	0	14	0	15	15
Vagina	Projected in 2042	172	168	167	167	167	167	166	166	165	166	165	164
	Prevented in 2042	0	4	5	5	5	5	6	6	7	6	7	8
	Prevented 2018–2042	0	52	55	55	58	57	61	60	64	62	66	69
Vulva	Projected in 2042	987	964	960	961	956	957	953	953	949	951	947	945
	Prevented in 2042	0	23	27	27	31	30	34	34	38	36	40	42
	Prevented 2018–2042	0	296	313	309	327	323	341	340	358	350	368	384
Total, women	Projected in 2042	3873	3613	3568	3578	3533	3544	3499	3501	3456	3475	3430	3406
	Prevented in 2042	0	260	305	295	340	330	374	372	417	399	443	468
	Prevented 2018–2042	0	3291	3480	3436	3625	3582	3771	3762	3951	3873	4062	4174
Total, men	Projected in 2042	605	605	603	605	603	605	603	605	603	605	602	602
	Prevented in 2042	0	0	2	0	2	0	2	0	2	0	2	3
	Prevented 2018–2042	0	0	21	0	22	0	24	0	25	0	27	40
Total, both	Projected in 2042	4478	4218	4171	4183	4136	4149	4102	4106	4059	4080	4032	4007
	Prevented in 2042	0	260	307	295	342	330	376	372	419	399	446	471
	Prevented 2018–2042	0	3291	3501	3436	3648	3582	3794	3762	3977	3873	4089	4213

[a] We did not round numbers when performing the analysis and hence some figures do not add up.
[b] The direct effects of 80% vaccine coverage among boys was not modeled.
[c] Since cancer incidence was projected to only 2042, the first vaccinated cohort of girls vaccinated in 2008 at ages 10–14 were aged 40–44 in 2042 meaning that only cancer incidence among those up to age 45 could be impacted by vaccination.
[d] Direct effects of 0.0 among girls and boys assume that the HPV vaccination was never administered at any point in time in Canada.
[e] We estimated that 49.4% of anal cancers occur among men who have sex with men and hence are not impacted by herd effects of girls only vaccination.

we multiplied the following parameters: (1) proportion of cancer attributable to high-risk HPV types for anogenital cancers (ranging from 39.4% for penile cancer to 100.0% for cervical cancer – Table 1) and to HPV16 for head and neck cancers (ranging from 8.2% for oral cavity cancer to 60.2% for oropharyngeal cancer), (2) level of direct (40.0%–80.0%) or herd (36.0%–100.0%) vaccine coverage effects, (3) level of protection offered by the vaccines (66.0%–97.7%), and (4) vaccine efficacy (95.0%). The resulting proportion was then multiplied by the projected number of cancers to calculate the number of preventable cancers.

2.6. Cancer incidence

Supplementary Table 1 describes the modeling approach to estimate future cancer incidence (2018–2042) for each cancer. The projected number of cancers was estimated using three methods. The first method involved fitting different models with the 'Canproj' R package; this process is described in detail elsewhere (Poirier et al., 2019). The second involved applying a proportion to the Canproj projected cancer incidence to obtain the number of incident cancers for specific subsites. For example, this approach was utilized to determine the proportion of tongue cancer that is expected to be from the base of tongue and the proportion of stomach cancer that is expected to be from the non-cardia part of the stomach (Supplementary Table 1). Cancer incidence data for rare or subsite cancers were only available for two age groups (< 50 and ≥ 50 years). To approximate the number of cancers occurring in five-year age groups, we partitioned the counts in these two age groups by the five-year age distributions from other related cancers. Specifically, the cervical cancer five-year age distribution within the < 50 and ≥ 50 age groups was used to partition vaginal and vulvar cancers, and the tongue cancer five-year age distribution was used to partition tonsillar cancer, thereby allowing us to assess the impact of HPV vaccination on cancer incidence. As herd effects from girls-only vaccination do not confer protection among men who have sex with men (MSM), we estimated the proportion of anal cancers occurring among MSM. We calculated a proportion of 49.4% of anal cancers attributable to MSM by utilizing a RR of 17.3 for the association between sexual orientation and anal cancer and a 6.0% prevalence of MSM among those aged 15 to 44 in the United States (Chandra et al., 2011; Daling et al., 2004).

2.7. Statistical analysis

The calculation of attributable risks has been previously published (Volesky et al., 2019). Briefly, to estimate the proportion of cancer that is attributable to HPV, individual studies were pooled with a random effects model. A fixed effect model was used to produce a pooled measure of association between *H. pylori* and non-cardia gastric cancer. Meta-analyses were performed, and figures were produced in Stata v14 (StataCorp., College Station, TX, USA). R software (version 3.4.1) was used to calculate the future preventable burden of HBV, HCV, and *H.*

Table 5
Projected number of head and neck cancers and the number that could be prevented with variations in school-based HPV vaccine coverage in Canada[a,b,c].

Cancer site, sex	Future burden measures	Lower (%)							Current (%)		Higher (%)		
	Effect:	Direct [d]	Direct	Herd	Direct	Herd	Direct	Herd	Direct	Herd	Direct	Herd	Herd
	Girls:	0.0	40.0	53.0	50.0	63.0	60.0	73.0	72.4	85.4	80.0	93.0	100.0
	Boys:	0.0	0.0	36.0	0.0	46.0	0.0	56.0	0.0	68.4	0.0	83.0	100.0
Oropharynx, men[e]	Projected in 2042	3469	3469	3363	3469	3360	3469	3356	3469	3352	3469	3348	3342
	Prevented in 2042	0	0	106	0	109	0	113	0	117	0	121	127
	Prevented 2018–2042	0	0	742	0	760	0	778	0	800	0	826	857
Oropharynx, women[e]	Projected in 2042	914	894	892	892	890	891	889	889	887	888	885	884
	Prevented in 2042	0	20	23	22	24	24	26	26	28	27	29	30
	Prevented 2018–2042	0	186	193	192	199	198	205	205	213	210	217	227
Oropharynx, both[e]	Projected in 2042	4383	4363	4255	4361	4250	4360	4245	4358	4239	4357	4233	4226
	Prevented in 2042	0	20	129	22	134	24	138	26	145	27	151	157
	Prevented 2018–2042	0	186	935	192	959	198	983	205	1013	210	1044	1084
Oral cavity, men	Projected in 2042	761	761	760	761	760	761	759	761	759	761	759	759
	Prevented in 2042	0	0	2	0	2	0	2	0	2	0	2	2
	Prevented 2018–2042	0	0	16	0	16	0	17	0	17	0	18	19
Oral cavity, women	Projected in 2042	858	856	856	856	856	856	856	856	855	855	855	855
	Prevented in 2042	0	2	2	2	2	2	3	3	3	3	3	3
	Prevented 2018–2042	0	22	24	23	24	24	25	25	26	26	27	28
Oral cavity, both	Projected in 2042	1619	1617	1616	1617	1615	1617	1615	1617	1615	1617	1615	1614
	Prevented in 2042	0	2	4	2	4	2	4	3	5	3	5	5
	Prevented 2018–2042	0	22	40	23	41	24	42	25	44	26	45	46
Larynx, men	Projected in 2042	1230	1229	1229	1230	1228	1230	1228	1230	1228	1230	1228	1228
	Prevented in 2042	0	0	2	0	2	0	2	0	2	0	2	2
	Prevented 2018–2042	0	0	11	0	12	0	12	0	12	0	12	12
Larynx, women	Projected in 2042	187	186	186	186	186	186	186	186	186	186	186	186
	Prevented in 2042	0	0	0	0	0	0	0	0	0	0	0	1
	Prevented 2018–2042	0	4	4	4	4	4	5	5	5	5	5	5
Larynx, both	Projected in 2042	1417	1415	1415	1417	1415	1416	1415	1416	1415	1416	1415	1415
	Prevented in 2042	0	0	2	0	2	0	2	0	2	0	2	2
	Prevented 2018–2042	0	4	16	4	16	4	16	5	16	5	17	17
Total, men	Projected in 2042	5460	5459	5351	5460	5348	5460	5344	5460	5340	5460	5335	5330
	Prevented in 2042	0	0	109	0	113	0	116	0	120	0	125	131
	Prevented 2018–2042	0	0	769	0	788	0	806	0	829	0	856	888
Total, women	Projected in 2042	1959	1937	1934	1935	1932	1933	1931	1931	1928	1929	1927	1926
	Prevented in 2042	0	23	25	25	27	26	29	29	31	30	33	34
	Prevented 2018–2042	0	212	221	219	228	226	235	235	244	240	249	260
Total, both	Projected in 2042	7420	7395	7285	7395	7280	7393	7275	7391	7268	7390	7262	7255
	Prevented in 2042	0	23	135	25	140	26	145	29	151	30	158	165
	Prevented 2018–2042	0	212	990	219	1016	226	1041	235	1073	240	1105	1148

[a] We did not round numbers when performing the analysis and hence some figures do not add up.
[b] Direct effects of 80% vaccine coverage among boys was not modeled.
[c] Since cancer incidence was projected to only 2042, the first vaccinated cohort of girls vaccinated in 2008 at ages 10–14 were aged 40–44 in 2042 meaning that only cancer incidence among those up to age 45 could be impacted by vaccination.
[d] Direct effects of 0.0 among girls and boys assume that the HPV vaccination was never administered at any point in time in Canada.
[e] Included the base of the tongue and tonsils.

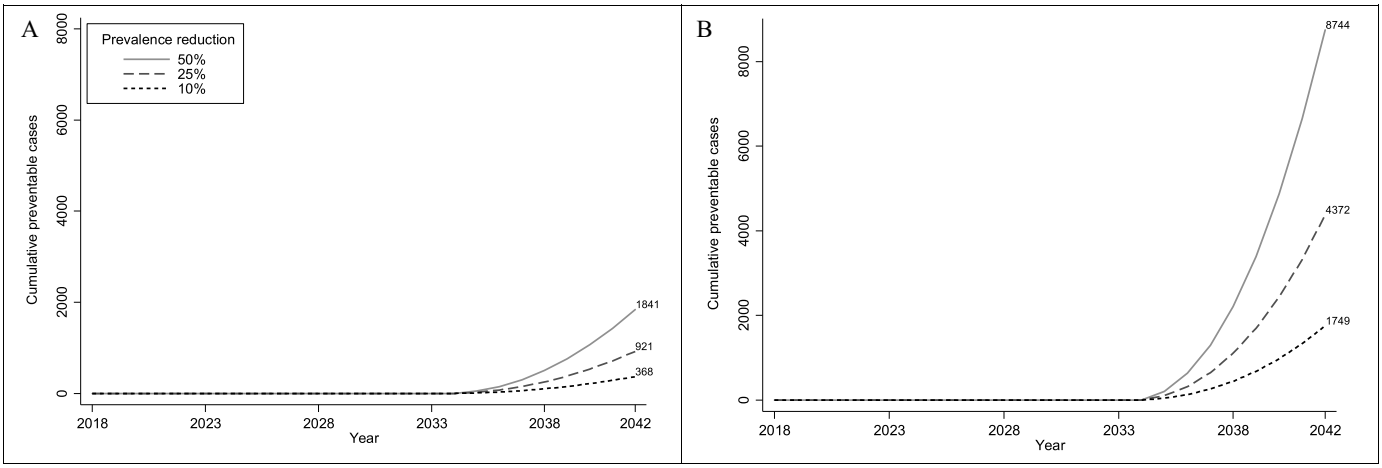


Fig. 1. Projected cumulative preventable cases attributable to hepatitis B and C viruses (A) and *Helicobacter pylori* (B) by applying counterfactual prevalence reductions.

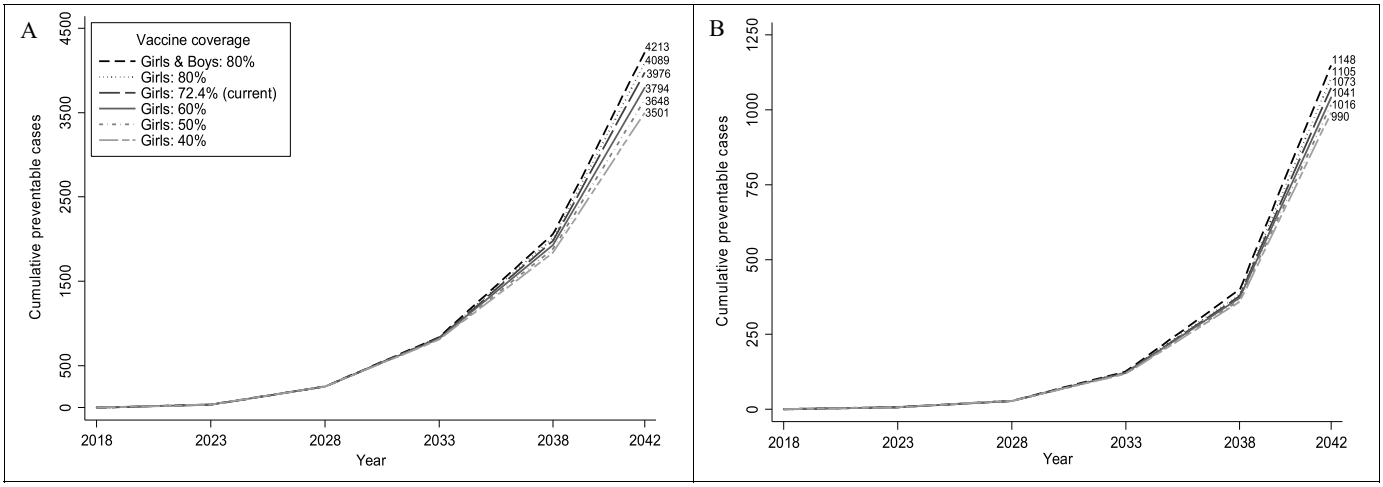


Fig. 2. Projected cumulative preventable anogenital cancers (A) and head and neck cancers (B) attributable to human papillomavirus by applying school-based HPV vaccine coverage counterfactuals[a,b].

[a]The vaccine coverage level refers to the percent of those aged 10–14 receiving the HPV vaccine.

[b]We modeled the herd effects of vaccine coverage (e.g. 40% coverage of girls produces 53% coverage of girls and 36% coverage of boys).

pylori associated cancers (R Foundation for Statistical Computing [Internet], 2017) and an electronic spreadsheet was used to estimate the future preventable burden of HPV associated cancers.

Ethics approval was granted for this project by the Health Research Ethics Board of Alberta - Cancer Committee (HREBA.CC-14-0220_REN4) and McGill University exempted this study from Research Ethics Board review.

3. Results

A 50% reduction in HBV, HCV, *H. pylori* prevalence and 80% HPV vaccine coverage of girls and boys in 2018 resulted in an estimated 15,946 cancers that could be prevented from 2018 to 2042 (Tables 2–5). Figs. 1 and 2 demonstrate the cumulative increase in the number of preventable cases over time, and for HBV, HCV and *H. pylori* after a latency period. A 50% reduction in the prevalence of HBV, HCV and *H. pylori* and 80% HPV coverage among girls and boys, could prevent an estimated 1.0% of all cancers diagnosed among men and 0.9% diagnosed among women in 2042 (data not shown).

3.1. Hepatitis B and C viruses

The future prevalence of HBV remained constant, however, that of HCV was projected as steadily decreasing to 2042. A 50% reduction in the prevalence of HBV and HCV would result in slightly fewer projected hepatocellular carcinoma cases in 2042; from 3358 to 3210 for HBV and from 3358 to 3106 for HCV (Table 2). Cumulatively from 2018 to 2042, a 10% reduction in the prevalence of HBV and HCV would prevent 356 hepatocellular carcinomas as compared to a 50% prevalence reduction that would prevent 1782 hepatocellular carcinomas.

3.2. Helicobacter pylori

A 50% prevalence reduction in *H. pylori* would lead to fewer projected non-cardia gastric cancers (3579 cases) in 2042 compared to no change in prevalence (5097 cases); and fewer gastric mucosa-associated lymphoid tissue (MALT) lymphomas, from 2403 to 1822 (Table 3). Cumulatively from 2018 to 2042, a 10% reduction in the prevalence of *H. pylori* would prevent 1749 non-cardia gastric cancers and gastric MALT lymphoma cases as compared to a 50% prevalence reduction that would result in 8744 fewer cases.

3.3. Human papillomavirus

If the estimated current Canada-wide HPV vaccine coverage of girls continued (72.4% direct coverage, but due to herd effects becoming equivalent to 85.4% coverage in girls and 68.4% in boys), an estimated total of 3976 anogenital cancers could be prevented from 2018 to 2042 (Table 4). The majority (85.4%) of these preventable cases were cervical cancers, and virtually all preventable cases occurred among women (99.4%). In contrast, continuation of current HPV vaccine coverage could prevent more head and neck cancers among men (829 cases) than women (244 cases) from 2018 to 2042 (Table 5). Among all HPV-caused cancers, 80% vaccine coverage of girls and boys could prevent 4434 cancers among women and 928 among men by 2042 (Tables 4 and 5) in those less than age 45.

4. Discussion

4.1. Hepatitis B and C viruses

The World Health Organization developed a global strategy to eliminate viral hepatitis with a focus on HBV and HCV by 2030 (World Health Organization, 2016); Canada is a signatory to this strategy. For HBV, the major prevention measure is vaccination, which began as early as 1982 in Canada (Government of Canada, 2017). The Canadian government encourages health care providers to assess HBV status and immunize persons immigrating to Canada (Government of Canada, 2017), although this immunization does not appear to be systematic. The future incidence of hepatocellular carcinoma would be impacted by school- or infant- based universal immunization making a 50% reduction in the prevalence plausible. Approximately 12% of hepatocellular carcinoma cases could be prevented in 2042 with a 50% reduction in the prevalence of the hepatitis viruses in 2018; a 10% reduction would prevent only 2.4% of hepatocellular carcinoma cases in 2042. However, incorporating a 15-year latency for HBV and HCV provided only a 10-year window (from 2032 to 2042) where cancer incidence could be changed by prevalence reductions.

4.2. Helicobacter pylori

H. pylori was the infectious agent responsible for the most preventable cancer cases from 2018 to 2042 (8744 cancers with a 50% prevalence reduction). Although *H. pylori* is associated with a prolonged latency thereby expanding the opportunity to detect and deliver

quadruple antibiotic therapy, there are challenges around determining who needs to be screened and concerns over increasing antibiotic resistance (Fallone et al., 2016). A 50% prevalence reduction may be more aspirational than attainable; however, the more achievable 25% prevalence reduction could prevent > 4000 cancers from 2018 to 2042. When we projected the future prevalence of *H. pylori*, we assumed a constant trend. Nonetheless, a decreasing trend in its prevalence would have resulted in fewer prevented cases, and an increasing trend would have resulted in more.

4.3. Human papillomavirus

A 40% vaccination coverage of girls (herd effects lead to 53% coverage equivalents among girls and 36% among boys) achieved a notable number of preventable cases, with 4491 potentially preventable cancers from 2018 to 2042. Since we used a birth cohort approach, the first two five-year cohorts were vaccinated prior to the application of counterfactual vaccine coverage in 2018, and thus the counterfactuals' impact on cancer incidence in these two cohorts was not modeled.

By projecting cancer incidence to only 2042, the first cohort of girls vaccinated in 2008 at ages 10 to 14 was then aged 40 to 44 in 2042, therefore only cancer incidence among individuals up to age 45 could be impacted. For boys, this constraint was even more pronounced as the vaccine was assumed to have been delivered starting in 2018. This restriction greatly influenced our results since only the first two cohorts could be followed to ages 35 to 44, whereas the remaining cohorts could only be followed to ages 30 to 34. Specifically, the impact of HPV vaccination counterfactuals was confined to cancers occurring in individuals under age 35 in 2042 and therefore differences between the counterfactual interventions are minimized as these only apply to younger cohorts. The impact of HPV vaccine coverage was limited to cancers occurring among individuals less than age 45, yet the majority of HPV-related cancers occurred in individuals over age 45. Hence, our analysis provided a short-term assessment of the impact of school-based vaccination on cancer incidence among young Canadians.

Modeling the impact of HPV vaccine coverage counterfactuals involved several assumptions. First, the estimated herd effects relied on informed assumptions about the level of protection (40% and 80% direct coverage) among non-vaccinated individuals (Brisson et al., 2016) but had to be interpolated for other coverage levels modeled here (i.e. 50%, 60%, 72.4%). Second, we used a more conservative approach to estimate current country-wide vaccine coverage by utilizing data on the completion of recommended number of doses; yet, one dose has been shown to offer considerable protection against HPV-related diseases (Kreimer et al., 2015). Third, we assumed that the vaccine confers long-term protection (up to 30 years in our calculations) against the HPV types it protects against.

There are several limitations of our analysis. First, we did not account for immigration in our calculations; for example, new arrivals not vaccinated through school-based or catch-up vaccination programs were not accounted for by the counterfactuals and they have a greater risk of developing HPV-associated cancers than the remaining Canadian population; however, herd effects are anticipated to minimize this concern. Second, although our estimate of country-wide vaccination was conservative (72.4%), there is substantial variation in the level of HPV vaccine coverage, hence some Canadian jurisdictions might not realize the reductions in cancer incidence that are possible with the counterfactual coverage levels. For example, receiving the recommended number of vaccine doses ranges from approximately 50% in Nunavut to 90% in Newfoundland and Labrador (Shapiro et al., 2017). Conservatively, the impact of catch-up vaccination was not modeled, yet it would result in more preventable cancers in the future. Third, improvements in cervical cancer screening technology coupled with vaccination coverage are likely to result in improved and more efficient cervical cancer prevention in the future, potentially leading to elimination of this disease (El-Zein et al., 2016; Franco, 2017). Finally, we focused our analysis on the four infections that cause the most cancers in Canada and for which there are proven prevention strategies; however, other infections such as Epstein-Barr virus and human immunodeficiency virus also cause cancer and a reduction in their prevalence could lessen the future infection-associated cancer burden.

4.4. Implications for cancer prevention

With an aging population, the future burden of cancer in Canada is expected to substantially increase to 2032 (Canadian Cancer Society's Advisory Committee on Cancer Statistics, 2015). Changes in major cancer risk factors such as infections will have varying impacts on the future burden of cancer; we identified the impact of four preventable and/or treatable infections on the future cancer burden. Even the short-term view presented here reveals that different interventions have differing impacts on future incidence.

5. Conclusion

By modeling the impact of 10%, 25%, and 50% relative reductions in the prevalence of infections – HBV, HCV, and *H. pylori* – we estimated that > 10,000 cancers could be prevented from 2018 to 2042 with a 50% prevalence reduction. The impact of 80% school-based HPV vaccine coverage among girls and boys would potentially prevent 5360 cancer cases from 2018 to 2042. Despite only capturing the impact of school-based HPV vaccination on cancers occurring among those less than age 45, our results indicate that increases in HPV coverage can result in meaningful decreases in HPV-related cancer incidence. With Canada's current cancer prevention resources, there is a substantial opportunity to reduce the future infection-associated cancer burden.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ypmed.2019.04.006.

Acknowledgements

Karena D. Volesky is supported by a doctoral award from the Fonds de recherche du Québec - Santé and Fondation des étoiles. Darren Brenner holds a Canadian Cancer Society Prevention Capacity Development Award (#703917). Christine Friedenreich was supported by an Alberta Innovates Health Senior Scholar Award and by the Alberta Cancer Foundation Weekend to End Women's Cancers Breast Cancer Chair. We are grateful to Talía Malagón for her modeling advice, and to Shemay Lee for assisting us with data management.

Competing interests

None.

Disclosure

E.L.F. has served as occasional consultant to companies involved with HPV diagnostics and vaccination (Merck, GSK, Roche, and BD). His institution has received grants from Merck and Roche to supplement investigator-initiated studies that he leads at McGill University.

E.L.F. is Editor-in-Chief at Preventive Medicine and K.D.V. is an Assistant Editor at Preventive Medicine. The process of soliciting the special issue, sending out manuscripts for review, the peer-review process and editorial decision making was conducted entirely outside of the Preventive Medicine online system (for which E.L.F. and K.D.V. have access to through their regular Preventive Medicine duties).

Funding sources

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

Role of funding source

The sponsor had no role in the study design, data collection, data analysis or interpretation, writing of the manuscript or in the decision to submit this manuscript for publication.

References

Bishop, J.A., Ma, X.J., Wang, H., Luo, Y., Illei, P.B., Begum, S., et al., 2012. Detection of transcriptionally active high-risk HPV in patients with head and neck squamous cell carcinoma as visualized by a novel E6/E7 mRNA in situ hybridization method. Am. J. Surg. Pathol. 36 (12), 1874–1882.

Bosch, F.X., Broker, T.R., Forman, D., Moscicki, A.B., Gillison, M.L., Doorbar, J., et al., 2013. Comprehensive control of human papillomavirus infections and related diseases. Vaccine. 31 (Suppl. 6), G1–31.

Brenner, D.R., Poirier, A.E., Walter, S.D., King, W.D., Franco, E.L., Demers, P.A., et al., 2018. Estimating the current and future cancer burden in Canada: methodological framework of the Canadian population attributable risk of cancer (ComPARe) study. BMJ Open 8 (7), e022378.

Brisson, M., Benard, E., Drolet, M., Bogaards, J.A., Baussano, I., Vanska, S., et al., 2016. Population-level impact, herd immunity, and elimination after human papillomavirus vaccination: a systematic review and meta-analysis of predictions from transmission-dynamic models. Lancet Public Health 1 (1), e8–e17.

Canadian Cancer Society's Advisory Committee on Cancer Statistics, 2015. Canadian Cancer Statistics. Canadian Cancer Society, Toronto, ON.

Centers for Disease Control and Prevention, 2001. In: Department of Health and Human Services (Ed.), NHANES 1999–2000 Laboratory Data.

Centers for Disease Control and Prevention, 2009. In: Department of Health and Human Services (Ed.), NHANES 2007–2008 Laboratory Data.

Centers for Disease Control and Prevention, 2011. In: Department of Health and Human Services (Ed.), NHANES 2009–2010 Laboratory Data.

Chandra, A., Mosher, W.D., Copen, C., Sionean, C., 2011. Sexual behavior, sexual attraction, and sexual identity in the United States: data from the 2006–2008 National Survey of Family Growth. Natl. Health Stat Rep. (36), 1–36.

Cheung, J., Goodman, K.J., Girgis, S., Bailey, R., Morse, J., Fedorak, R.N., et al., 2014. Disease manifestations of *Helicobacter pylori* infection in Arctic Canada: using epidemiology to address community concerns. BMJ Open 4 (1), e003689.

Daling, J.R., Madeleine, M.M., Johnson, L.G., Schwartz, S.M., Shera, K.A., Wurscher, M.A., et al., 2004. Human papillomavirus, smoking, and sexual practices in the etiology of anal cancer. Cancer. 101 (2), 270–280.

Deeks, S.L., Tunis, M.C., Ismail, S., National Advisory Committee on I, 2017. Summary of the NACI update on the recommended use of human papillomavirus (HPV) vaccine: nine-valent HPV vaccine two-dose immunization schedule and the use of HPV vaccines in immunocompromised populations. Can. Commun. Dis. Rep. 43 (6), 138–142.

Drolet, M., Benard, E., Boily, M.C., Ali, H., Baandrup, L., Bauer, H., et al., 2015. Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. Lancet Infect. Dis. 15 (5), 565–580.

El-Serag, H.B., 2012. Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology. 142 (6), 1264–73 e1.

El-Zein, M., Richardson, L., Franco, E.L., 2016. Cervical cancer screening of HPV vaccinated populations: cytology, molecular testing, both or none. J. Clin. Virol. 76 (Suppl. 1) (S62-S8).

Fallone, C.A., Chiba, N., van Zanten, S.V., Fischbach, L., Gisbert, J.P., Hunt, R.H., et al., 2016. The Toronto consensus for the treatment of *Helicobacter pylori* infection in adults. Gastroenterology. 151 (1), 51–69 e14.

Ford, A.C., Forman, D., Hunt, R.H., Yuan, Y., Moayyedi, P., 2014. Helicobacter pylori eradication therapy to prevent gastric cancer in healthy asymptomatic infected individuals: systematic review and meta-analysis of randomised controlled trials. BMJ. 348, g3174.

Franco, E., 2017. Guest editorial: outlook on the prevention and control of HPV-associated cancers in Canada. In: Canadian Cancer Society's Advisory Committee on Cancer Statistics (Ed.), Canadian Cancer Statistics 2016. Canadian Cancer Society, Toronto, ON, pp. 99–100.

Government of Canada, 2017. Canadian immunization guide: part 4 - active vaccines. Available from: https://www.canada.ca/en/public-health/services/publications/healthy-living/canadian-immunization-guide-part-4-active-vaccines/page-7-hepatitis-b-vaccine.html.

Future II Study Group, 2007. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. N. Engl. J. Med. 356 (19), 1915–1927.

Hanley, S.J., Yoshioka, E., Ito, Y., Kishi, R., 2015. HPV vaccination crisis in Japan. Lancet. 385 (9987), 2571.

Huh, W.K., Joura, E.A., Giuliano, A.R., Iversen, O.E., de Andrade, R.P., Ault, K.A., et al., 2017. Final efficacy, immunogenicity, and safety analyses of a nine-valent human papillomavirus vaccine in women aged 16–26 years: a randomised, double-blind trial. Lancet. 390 (10108), 2143–2159.

Kreimer, A.R., Struyf, F., Del Rosario-Raymundo, M.R., Hildesheim, A., Skinner, S.R., Wacholder, S., et al., 2015. Efficacy of fewer than three doses of an HPV-16/18 AS04-adjuvanted vaccine: combined analysis of data from the Costa Rica vaccine and PATRICIA trials. Lancet Oncol. 16 (7), 775–786.

Lingala, S., Ghany, M.G., 2015. Natural history of hepatitis C. Gastroenterol. Clin. N. Am. 44 (4), 717–734.

de Martel, C., Plummer, M., Vignat, J., Franceschi, S., 2017. Worldwide burden of cancer attributable to HPV by site, country and HPV type. Int. J. Cancer 141, 664–670.

Morgenstern, H., Bursic, E.S., 1982. A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. J. Community Health 7 (4), 292–309.

Naja, F., Kreiger, N., Sullivan, T., 2007. Helicobacter pylori infection in Ontario: prevalence and risk factors. Can. J. Gastroenterol. 21 (8), 501–506.

Plummer, M., de Martel, C., Vigant, J., Ferlay, J., Bray, F., Franceschi, A., 2016. Global burden of cancer attributable to infections in 2012: a synthetic analysis. Lancet 4 (9) (e609-e16).

Poirier, A.E., Ruan, R., Walter, S.D., Franco, F.L., King, W.D., Villeneuve, P., et al., 2019. The Future Burden of Cancer in Canada: Long-term Cancer Incidence Projections 2013–2042. Cancer Epidemiol. 59, 199–207.

Public Health Agency of Canada, 2011. Brief report: hepatitis B infection in Canada. Available from. http://www.phac-aspc.gc.ca/id-mi/pdf/hepB-eng.pdf.

R: A language and environment for statistical computing. R Foundation for Statistical Computing Internet. Available from. https://www.R-project.org/.

Remis, R., 2010. Modelling the incidence of prevalence of hepatitis C infection and its sequelae in Canada, 2007. In: Public Health Agency of Canada.

Rietbergen, M.M., Leemans, C.R., Bloemena, E., Heideman, D.A., Braakhuis, B.J., Hesselink, A.T., et al., 2013. Increasing prevalence rates of HPV attributable oropharyngeal squamous cell carcinomas in the Netherlands as assessed by a validated test algorithm. Int. J. Cancer 132 (7), 1565–1571.

Saraiya, M., Unger, E.R., Thompson, T.D., Lynch, C.F., Hernandez, B.Y., Lyu, C.W., et al., 2015. US assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines. J. Natl. Cancer Inst. 107 (6), djv086.

Sethi, A., Chaudhuri, M., Kelly, L., Hopman, W., 2013. Prevalence of *Helicobacter pylori* in a first nations population in northwestern Ontario. Can. Fam. Physician 59 (4), e182–e187.

Shapiro, G.K., Guichon, J., Kelaher, M., 2017. Canadian school-based HPV vaccine programs and policy considerations. Vaccine. 35 (42), 5700–5707.

Statens Serum Institut, 2017. Human papillomavirus-vaccine (HPV) færdigvaccineret - vaccinationstilslutning. Available from. http://www.ssi.dk/Smitteberedskab/Sygdomsovervaagning.aspx.

Statistics Canada, 2010. Canadian Health Measures Survey (CHMS) Data User Guide: Cycle 1.

Statistics Canada Canadian Health Measures Survey (CHMS). Available from. http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5071.

Stein, A.P., Saha, S., Yu, M., Kimple, R.J., Lambert, P.F., 2014. Prevalence of human papillomavirus in oropharyngeal squamous cell carcinoma in the United States across time. Chem. Res. Toxicol. 27 (4), 462–469.

Trubnikov, M., Yan, P., Archibald, C., 2014. Estimated prevalence of hepatitis C virus infection in Canada, 2011. Can. Commun. Dis. Rep. 40 (19), 429–436.

Van de Velde, N., Boily, M.C., Drolet, M., Franco, E.L., Mayrand, M.H., Kliewer, E.V., et al., 2012. Population-level impact of the bivalent, quadrivalent, and nonavalent human papillomavirus vaccines: a model-based analysis. J. Natl. Cancer Inst. 104 (22), 1712–1723.

Volesky, K.D., El-Zein, M., Franco, E.L., Brenner, D.R., Friedenreich, C.M., Ruan, Y., 2019. Cancers Attributable to Infections in Canada. 122. pp. 109–117.

World Health Organization, 2016. Global Health Sector Strategy on Viral Hepatitis 2016–2021. Geneva, Switzerland. .