Global burden of cancers attributable to infections in 2008: a review and synthetic analysis

Catherine de Martel, Jacques Ferlay, Silvia Franceschi, Jérôme Vignat, Freddie Bray, David Forman, Martyn Plummer

Summary

Background Infections with certain viruses, bacteria, and parasites have been identified as strong risk factors for specific cancers. An update of their respective contribution to the global burden of cancer is warranted.

Methods We considered infectious agents classified as carcinogenic to humans by the International Agency for Research on Cancer. We calculated their population attributable fraction worldwide and in eight geographical regions, using statistics on estimated cancer incidence in 2008. When associations were very strong, calculations were based on the prevalence of infection in cancer cases rather than in the general population. Estimates of infection prevalence and relative risk were extracted from published data.

Findings Of the 12·7 million new cancer cases that occurred in 2008, the population attributable fraction (PAF) for infectious agents was 16·1%, meaning that around 2 million new cancer cases were attributable to infections. This fraction was higher in less developed countries (22·9%) than in more developed countries (7·4%), and varied from 3·3% in Australia and New Zealand to 32·7% in sub-Saharan Africa. *Helicobacter pylori*, hepatitis B and C viruses, and human papillomaviruses were responsible for 1·9 million cases, mainly gastric, liver, and cervix uteri cancers. In women, cervix uteri cancer accounted for about half of the infection-related burden of cancer; in men, liver and gastric cancers accounted for more than 80%. Around 30% of infection-attributable cases occur in people younger than 50 years.

Interpretation Around 2 million cancer cases each year are caused by infectious agents. Application of existing public health methods for infection prevention, such as vaccination, safer injection practice, or antimicrobial treatments, could have a substantial effect on the future burden of cancer worldwide.

Funding Fondation Innovations en Infectiologie (FINOVI) and the Bill & Melinda Gates Foundation (BMGF).

Lancet Oncol 2012; 13: 607–15

Published Online
May 9, 2012
DOI:10.1016/S1470-2045(12)70137-7

See Comment 564

International Agency for Research on Cancer, Lyon, France (C de Martel MD, J Ferlay ME, S Franceschi MD, J Vignat MSc, F Bray PhD, D Forman PhD, M Plummer PhD)

Correspondence to:
Dr Martyn Plummer, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon cedex 08, France
plummerm@iarc.fr

Introduction

Infection is recognised as a major cause of cancer worldwide. Prevention and treatment of infectious agents have already had a substantial effect on cancer prevention.[1] A useful statistic to quantify this effect is the population attributable fraction (PAF), defined as the proportion of new cancer cases in a specific population that would have been prevented by a hypothetical intervention on a specific exposure. For infectious agents classified as carcinogenic to humans,[2] we calculated the PAF worldwide and in eight regions, using GLOBOCAN statistics on estimated cancer incidence in 2008.[3] Similar calculations have been done for cancer incidence data from 1990[4] and 2002.[5] In this report, we substantially revised the methods to reduce uncertainties and biases resulting from lack of data on population-specific and age-specific infection prevalence. We also discuss a framework for calculating global attributable fractions that might be applied to other causes of cancer. Some physical or chemical carcinogens act synergistically with infectious agents to cause cancers; in these cases, the attributable fractions can add to more than 100%. We report the attributable fractions of infectious agents but do not report the contribution of any non-infectious cofactor.

Methods

Infectious agents

In February, 2009, an expert working group reviewed infectious agents that have been classified as carcinogenic to humans by the International Agency for Research on Cancer (IARC) Monographs programme.[6] Panel 1 shows these agents and their associated cancers, namely *Helicobacter pylori (H pylori)*, hepatitis B virus (HBV), hepatitis C virus (HCV), *Opisthorchis viverrini*, *Clonorchis sinensis*, human papillomavirus (HPV), Epstein-Barr virus (EBV), human T-cell lymphotropic virus type 1 (HTLV-1), human herpes virus type 8 (HHV-8; also known as Kaposi's sarcoma herpes virus), and *Schistosoma haematobium*. Other cancer sites and infectious agents for which the evidence of carcinogenicity is weaker, or that have not been evaluated by the IARC Monographs programme, are not considered. Since HIV causes cancer through immunosuppression, thereby enhancing the carcinogenic action of other viruses, it is considered a cofactor and a PAF is not separately calculated.[2] The appendix includes a short review of each infectious agent and its associated cancer sites.

See Online for appendix

Geographical areas

Global estimates of the number of cancer cases caused by infections were calculated for the eight geographical regions shown in figure 1; these are based on UN

geographical regions and were used by the GLOBOCAN 2008 project. Additional estimates were calculated for China, India, Japan, and Australia and New Zealand because of their large population sizes within their respective regions. For some infections, calculations were restricted to the corresponding endemic areas. Countries were classified by development status using UN definitions.[7] All countries in Europe and North America (as shown in figure 1), as well as Australia, New Zealand, and Japan, were considered more developed; all other countries were considered less developed.

Source of cancer incidence data

Estimates of the number of new cancer cases in 2008 were obtained from the GLOBOCAN 2008 report,[8] which provided age-specific and sex-specific incidence estimates for 27 cancers in 184 countries. Incidence estimates were directly available for seven of the infection-associated cancers considered in the present report (cancer of the liver, cervix, nasopharynx, bladder, Hodgkin's lymphoma, non-Hodgkin lymphoma, and Kaposi's sarcoma in sub-Saharan Africa). Some specific cancer subsites were included as part of broader categories in GLOBOCAN 2008, so incidence data were not readily available. Anal cancer (International Classification of Diseases code C21) was included in the colorectal cancer category (C18–C21). Oropharyngeal sites including tonsils and base of tongue (C01, C09, and C10) were included in the lip, oral cavity (C00–08), and other pharynx (C09, C10, C12–14) categories. Vulval, vaginal, and penile carcinoma, and Kaposi's sarcoma outside sub-Saharan Africa, were included in the category of other and unspecified cancers. For these cancers, we estimated the number of cases by multiplying the estimated number of cases in the broader GLOBOCAN 2008 category by the proportion of cancer subsites or subtypes from cancer registry data. These proportions were stratified by region, age, and sex, and, in general, they were derived from the same registry data used for the GLOBOCAN 2008 estimates. However, when these subtype proportions were based on microscopically verified cases (Burkitt's lymphoma, gastric lymphoma, and adult T-cell leukaemia) or specific anatomic location (non-cardia gastric cancers), only cancer registries with the highest standard of information were used.[9,10]

Panel 1:* Major cancer sites associated with group 1 infectious agents

- Stomach: *Helicobacter pylori*
- Liver: Hepatitis B virus, hepatitis C virus (HCV), *Opisthorchis viverrini, Clonorchis sinensis*
- Cervix uteri: Human papillomavirus (HPV) with or without HIV
- Anogenital (penile, vulva, vagina, anus): HPV with or without HIV
- Nasopharynx: Epstein-Barr virus (EBV)
- Oropharynx: HPV with or without tobacco or alcohol consumption
- Kaposi's sarcoma: Human herpes virus type 8 with or without HIV
- Non-Hodgkin lymphoma: *H pylori*, EBV with or without HIV, HCV, human T-cell lymphotropic virus type 1
- Hodgkin's lymphoma: EBV with or without HIV
- Bladder: *Schistosoma haematobium*

*Classified as carcinogenic to humans in International Agency for Research on Cancer Monograph 100B.[2]

Attributable-risk calculation

The number of new cancer cases attributable to each infection was calculated by multiplying incidence estimates by PAF. PAF is an estimate of the proportion of cases of a disease that could theoretically be avoided if exposure to a specific risk factor was modified or removed. This estimate relies on strong causal assumptions and a simplified statistical model. PAF combines the magnitude of effect of a risk factor with its distribution in the population. For a binary exposure, which is either present or absent, PAF can be calculated as

$$PAF = \frac{p(R-1)}{1+p(R-1)} \quad \text{(formula 1)}$$

where p is the prevalence of exposure in the general population and R is the relative risk associated with exposure. PAF can also be calculated retrospectively using the prevalence of cases (p_C):[11]

$$PAF = p_c \frac{(R-1)}{R} \quad \text{(formula 2)}$$

The quantity $(R-1)/R$ is also known as the attributable risk in the exposed (ARE). As R increases, ARE increases to a limit of 1, so that all cases among the exposed are attributed to the exposure. At this limit, PAF is equal to the prevalence in cases:

$$PAF = p_c \quad \text{(formula 3)}$$

We used this simplified formula when mechanistic knowledge strongly suggests that the presence of infection in a cancer is sufficient to infer that infection caused the cancer. When more than one infectious agent was associated with the same cancer site, we assumed that the infections do not interact, but represent alternate pathways to cancer. Thus, PAFs for different infections are additive.

Sources of infection prevalence and relative-risk data

Sources of data used for PAF calculations are summarised in table 1 and discussed in the appendix. For all cancers other than cholangiocarcinomas, we used retrospective calculation of PAF based on prevalence of infection in cancer cases (formula 2 or 3) rather than prevalence of infection in the population (formula 1). Large population-based surveys of infection prevalence are not always available and might not be representative of the population from which the cancer cases were

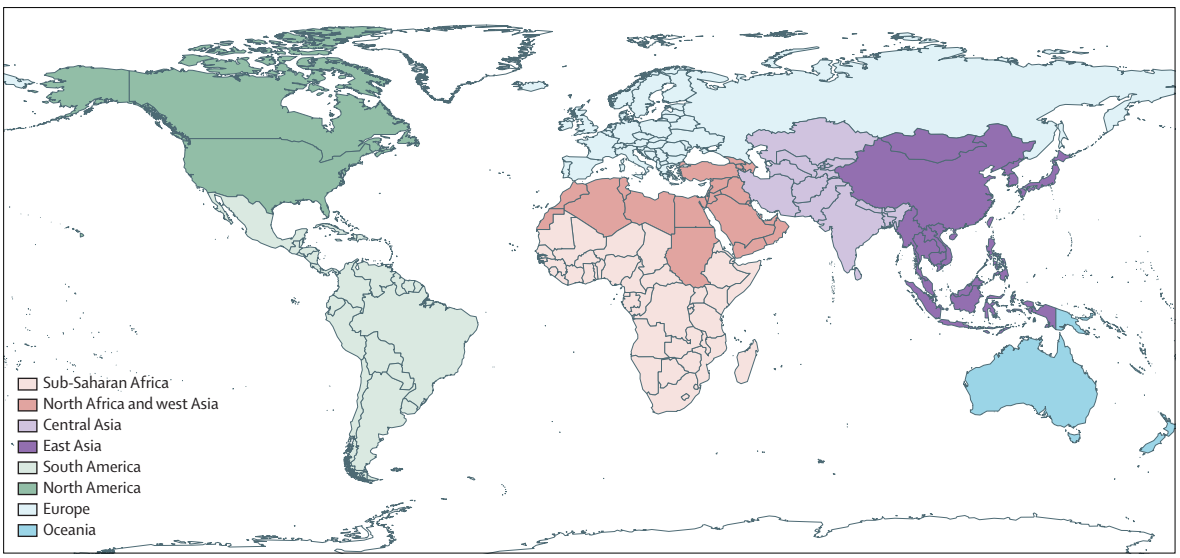

Figure 1: **Regions used to derive global estimates of the number of cancer cases caused by infection**

derived—they can often be focused on young and healthy subgroups such as blood donors, pregnant women, or military recruits. Population prevalence surveys might also be done for different age groups over different time periods, leading to difficulty calculating PAF for infections that show strong secular trends in prevalence such as *H pylori*, or that generally accumulate with age like HCV. Conversely, infection prevalence in cancer cases is usually straightforward to measure using molecular methods and can be assumed to directly represent the current population at risk.

For some infections (*H pylori*, HPV at anogenital sites, HHV-8, and HTLV-1), a global prevalence estimate was used because the literature showed no evidence of regional variations for prevalence of the infection in cases. For cancers that showed strong evidence of heterogeneity across countries, regional estimates of infection prevalence were calculated using one of two methods: geographical pooling or risk-based pooling. For geographical pooling, a weighted average of infection prevalence estimates for all countries in a region was calculated, where the weights were the product of the sample size of available studies of prevalence in cases and the number of incident cancer cases in each country given by GLOBOCAN 2008. This weighting scheme favours prevalence estimates from countries that make a larger contribution to the global cancer burden, while respecting the fact that larger prevalence surveys provide more statistical information. In geographical pooling, the underlying assumption is that countries within the same region have the same infection prevalence. For HBV and HCV in liver cancer, where geographical coverage of infection prevalence data was insufficient for some regions, risk-based pooling was used. The underlying assumption of risk-based pooling is that countries with similar incidence rates for a specific infection-associated cancer have similar infection prevalence. Countries were stratified by development status and a logistic regression model was fitted with infection prevalence as outcome and cancer incidence as predictor. Fitted values from this model were used to impute country-specific prevalence estimates for all countries, and these estimates were pooled by region and weighted by the number of incident cancers, as with geographical pooling.

Relative risk estimates for infection were extracted from studies reviewed in IARC Monograph 100B[2] or in formal meta-analyses. The association between an infection and its related cancer was assumed to be constant worldwide, and so a single relative risk estimate was used in the calculation of each PAF. SAS version 9.2 was used to compile the data and create the descriptive tables. R version 2.14.0 was used for specific calculations, such as risk-based pooling, and to create the figures. In the tables, estimated numbers of cases have been rounded to two significant figures to avoid spurious precision. In some cases, this creates small discrepancies with the displayed totals and percentages, which are based on the data before rounding.

Role of the funding source

The sponsors had no role in the study design, collection, analysis and interpretation of data, or writing of the report. CDM and MP had full access to all the data in the study and final responsibility for the decision to submit for publication.

Results

Table 2 shows the estimated number of cancer cases attributed to infection in 2008, in less developed and more developed regions. Of the estimated 12·7 million

new cancers worldwide, around 2 million were attributable to infections, of which 1·6 million (80%) occurred in less developed regions. HBV, HCV, HPV, and *H pylori* were together responsible for 1·9 million cases worldwide. Figure 2 shows the contribution of these infectious agents to cancer burden in less developed and more developed regions.

Table 3 shows a breakdown by geographical region of the number of new cancer cases and the number attributable to infection. Overall, 16·1% of cancer cases in 2008 were attributable to infection. The proportion was higher in less developed countries (22·9%) than in more developed countries (7·4%). The attributable fraction varied greatly between regions, from 3·3% in Australia and New Zealand to 32·7% in sub-Saharan Africa.

Table 4 shows a more detailed breakdown of attributable cancer cases, according to sex, age group, and development status of the country. Cervical cancer accounted for half of the attributable cases in women. The burden of gastric cancer and liver cancer was much higher in men than in women. The total number of cases attributable to infection was much the same in men and women. This similarity between sexes was noted across age groups, apart from in individuals younger than 40 years, where women had a higher burden of

	Relative risk (RR) estimate			Prevalence of infectious agent in cancer cases		
	Types of studies used for RR estimation	Laboratory method used for measurement of exposure	RR	Geographical area	Prevalence in cases (%)	Comments and strength of data
Helicobacter pylori						
Non-cardia gastric cancer† (C16.1–16.9)	Cohort with >10 years follow-up	ELISA in serum	5·9	World	90%	Data based on a pooled analysis of major prospective studies, all using ELISA serology.[12] Strong data
Non-Hodgkin lymphoma of gastric location† (MALT and DLBC) (C82–85, C96)	Cohort and case–control	ELISA in serum	7·2	World	86%	Consensus that nearly 100% of gastric MALT lymphomas are *H pylori*-related. Strong data No consensus on DLBC. Sparse data
Hepatitis viruses						
Liver cancer (C22)	Case–control (>10 cases)	HBV: HBsAg in serum HCV: ELISA in serum (first generation excluded)	23 17	Sub-Saharan Africa North Africa and west Asia South-central Asia India East Asia China Japan South America North America Europe Oceania	84% 82% 80% 79% 86% 86% 87% 82% 42% 48% 44%	For countries with no data, prevalence in cases imputed by liver cancer incidence and more developed or less developed status Data based on two large meta-analyses[13] and original data.[14] Strong data
Non-Hodgkin lymphomas (C82–85, C96)	Cohort and case–control	ELISA in serum (first generation excluded)	2·5	Southern Europe Japan and Korea Other more developed regions Less developed regions	18% 5% 10% 17%	Data based on one meta-analysis.[15] Limited data
Human papillomavirus (HPV; high-risk types)						
Cervix uteri carcinoma (C53)	Case–control	PCR in tumour tissue or cells	>100	World	100%	High-risk HPV types are considered a necessary cause of cervical cancer. Strong data
Penile carcinoma† (C60)	Case–control	PCR in tumour tissue	NR‡	World	50%	Assumption is that detection of high-risk HPV DNA in tumour tissue signifies cancer attributable to HPV Data based on one meta-analysis.[16] Limited data
Anal carcinoma† (C21)	Case–control	PCR in tumour tissue	NR‡	World	88%	Same assumption as for penile cancer Data based on one meta-analysis.[17] Strong data
Vulvar carcinoma† (C51)	Case–control	PCR in tumour tissue	NR‡	World	43%	Same assumption as for penile cancer Data based on one meta-analysis.[17] Limited data
Vaginal carcinoma† (C52)	Case–control	PCR in tumour tissue	NR‡	World	70%	Same assumption as for penile cancer Data based on one meta-analysis.[17] Limited data
Oropharynx† including tonsils and base of tongue (C01, C09–C10)	Case–control	PCR in tumour tissue with HPV E6 or E7 expression	NR‡	North America Northern and western Europe Eastern Europe Southern Europe Australia Japan Rest of world	56% 39% 38% 17% 45% 52% 13%	Few prevalence studies available for less developed regions Same assumption as for penile cancer, except for the difficulty separating strong effect of tobacco and alcohol. Limited data

(Continues on next page)

	Relative risk (RR) estimate			Prevalence of infectious agent in cancer cases		
	Types of studies used for RR estimation	Laboratory method used for measurement of exposure	RR	Geographical area	Prevalence in cases (%)	Comments and strength of data
(Continued from previous page)						
Epstein-Barr virus (EBV)						
Hodgkin's lymphoma (C81)	Cohort and case–control (>50 cases)	Detection of EBV gene products in tumour cells	>10	More developed regions Less developed regions	40% Children: 90% Adults: 60%	Prevalence varies by age, area, and histological subtype Shape of the age distribution curve varies by area and study period. Limited data
Burkitt's lymphoma† (C83.5)	Case–control and case series	EBV DNA in tumour cells	Sub-Saharan Africa: >20 USA and Europe: NR‡ Other regions: NR‡	Sub-Saharan Africa USA and Europe Other regions	100% 20% 30%	Sub-Saharan Africa and New Guinea are considered endemic areas for Burkitt's lymphoma in children (peak 4–7 years of age). Strong data Limited data for USA and Europe Sparse data for other regions
Nasopharyngeal carcinoma (C11)	Cohort and case–control	EBV DNA in tumour cells	>20	High-incidence and intermediate-incidence regions Low-incidence regions (see appendix for details)	100% 80%	High-incidence and intermediate-incidence: more than 95% of carcinomas are undifferentiated (type II), and nearly all are EBV-related. Strong data Low-incidence: only one cohort study and no case–control studies are available. The distribution between different types of carcinoma (types I, II, III) and the fraction attributable to EBV varies between countries and is largely unknown. Sparse data
Other non-Hodgkin lymphomas (C82–85, C96)	..	..	..	..	..	Due to the heterogeneity of this group, estimation of RR and prevalence is not possible from published data Data suggest that most non-Hodgkin lymphomas arising in people with HIV infection are causally related to EBV
Human herpes virus type 8 (HHV-8)						
Kaposi's sarcoma† (C46)	Cohort and case–control	HHV-8 DNA in tumour tissue	100	World	100%	Mostly (but not exclusively) seen in HIV-infected populations, particularly in Africa. Strong data
Human T-cell lymphotropic virus type 1 (HTLV-1)						
Adult T-cell leukaemia and lymphoma (ATL) (M9827, M9823)	Cohort and case–control	..	NR‡	World	100%	HTLV-1 seropositivity is required for diagnosis of ATL (HTLV-1 is a necessary cause). Strong data
Liver flukes (*Opisthorchis viverrini, Clonorchis sinensis*)						
Cholangiocarcinoma (C22.1)	Case–control (in endemic areas)	Various	7·7	Endemic areas in southeast Asia	NR‡	Estimates calculated prospectively using the same method as Parkin (2006).[5] Limited data
Schistosoma haematobium						
Bladder carcinoma (C67)	Case–control (in endemic areas)	Various	NR‡	Sub-Saharan Africa Egypt, Sudan, and Yemen	41% 42%	Assumption is that in endemic areas, all squamous-cell carcinomas could be attributed to *S haematobium*. Limited data

MALT=mucosa-associated lymphoid tissue. DLBC=diffuse large B-cell. HBV=hepatitis B virus. HBsAG=hepatitis B surface antigen. HCV=hepatitis C virus. NR=not relevant. *International Classification of Diseases (ICD)-10 code. †These subsites were not directly available in GLOBOCAN 2008; therefore, data from CI5-IX database were used to estimate corresponding incidence rates. ‡RRs were not available or not used in the calculation of PAF for these sites.

Table 1: **Methods for calculation of the population attributable fraction (PAF), by infectious agent and cancer site or subsite***

infection-related cancer, on account of cervical cancer, in less developed and more developed regions (figure 3).

Discussion

The analysis described in this report and in the appendix shows that infection is an important contributor to the global cancer burden, with 16·1% of cancers diagnosed in 2008 being attributable to infections, although the contribution due to infection varies widely from region to region. The estimated burden of cancer in 2008 attributable to infections is an update of previous estimates for 2002[5] and 1990.[4] Our estimates for 2008 are slightly lower than those for 2002, for global burden of cancer (16·1% *vs* 17·8%) and by development status (7·4% *vs* 7·7% for more developed regions; 22·9% *vs* 26·5% for less developed regions). Overall, the results are remarkably similar, in view of the change in methodology to incorporate retrospective calculation of PAF based on prevalence of infection in cancer cases. For the four main infections altogether, the relative contribution of HPV to cancer burden is similar in more developed and less developed areas. The contribution of

For more on the **CI5-IX database** see http://ci5.iarc.fr

	Less developed regions	More developed regions	World
Hepatitis B and C viruses	520 000 (32·0%)	80 000 (19·4%)	600 000 (29·5%)
Human papillomavirus	490 000 (30·2%)	120 000 (29·2%)	610 000 (30·0%)
Helicobacter pylori	470 000 (28·9%)	190 000 (46·2%)	660 000 (32·5%)
Epstein-Barr virus	96 000 (5·9%)	16 000 (3·9%)	110 000 (5·4%)
Human herpes virus type 8	39 000 (2·4%)	4100 (1·0%)	43 000 (2·1%)
Human T-cell lymphotropic virus type 1	660 (0·0%)	1500 (0·4%)	2100 (0·1%)
Opisthorchis viverrini and *Clonorchis sinensis*	2000 (0·1%)	0 (0·0%)	2000 (0·1%)
Schistosoma haematobium	6000 (0·4%)	0 (0·0%)	6000 (0·3%)
Total	1 600 000 (100·0%)	410 000 (100·0%)	2 000 000 (100·0%)

Data are number of new cancer cases attributed to a particular infectious agent (proportion of the total number of new cases attributed to infection that is attributable to a specific agent). *Numbers are rounded to two significant digits.

Table 2: **Number of new cancer cases* in 2008 attributable to infection, by infectious agent and development status**

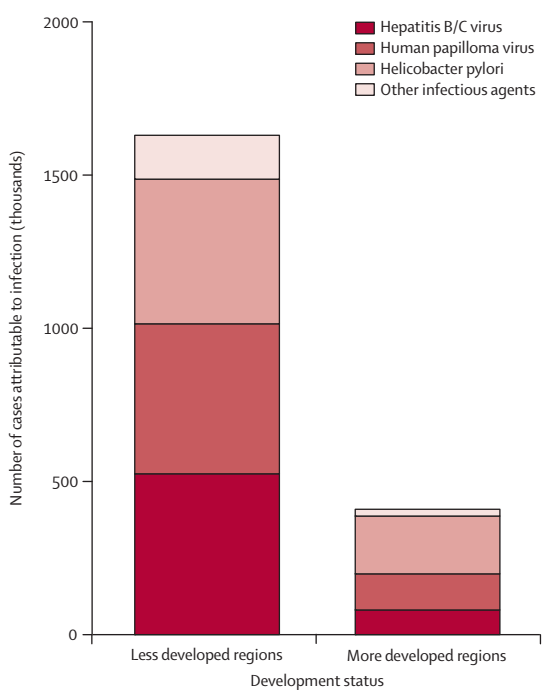

Figure 2: **Number of new cancer cases in 2008 attributable to infection, by infectious agent and development status**

H pylori is, however, proportionally larger in more developed countries, and that of HCV and HBV is larger in less developed countries.

Three recent studies have reported country-specific estimates of PAFs for infection-related cancer in the UK,[18] South Korea,[19] and China.[20] PAF estimates for China and South Korea were 25·9% and 21·2% respectively, in line with our estimates of 26·1% for China and 22·5% for east Asia. The estimate for the UK was 3·1%, which is lower than our regional estimate of 7·0% for Europe. We found similar estimates of PAF for North America (4·0%) and

	Number of new cases in 2008	Number attributable to infection	PAF (%)
Africa			
Sub-Saharan Africa	550 000	180 000	32·7%
North Africa and west Asia	390 000	49 000	12·7%
Asia			
India	950 000	200 000	20·8%
Other central Asia	470 000	81 000	17·0%
China	2 800 000	740 000	26·1%
Japan	620 000	120 000	19·2%
Other east Asia	1 000 000	230 000	22·5%
America			
South America†	910 000	150 000	17·0%
North America	1 600 000	63 000	4·0%
Europe	3 200 000	220 000	7·0%
Oceania			
Australia and New Zealand	130 000	4200	3·3%
Other Oceania	8800	1600	18·2%
More developed regions‡	5 600 000	410 000	7·4%
Less developed regions§	7 100 000	1 600 000	22·9%
World	12 700 000	2 000 000	16·1%

PAF=population attributable fraction. *Numbers are rounded to two significant digits. †Includes Mexico. ‡Total for Japan, North America, Europe, and Australia and New Zealand. §Total for all other regions.

Table 3: **Number of new cancer cases* in 2008 attributable to infectious agents, by geographical region**

Australia and New Zealand (3·3%), areas where infection prevalence is similar to the UK.

The main strengths of our approach are the use of the highest quality epidemiological evidence and incidence data available. The choice of infectious agents and cancer sites was taken from a review by an expert IARC working group.[2] We chose only agents classified as carcinogenic to humans by this group, and only cancer sites with sufficient evidence of association with infection (see appendix). Including more cancer sites—eg EBV in gastric cancer, HBV in non-Hodgkin lymphoma, or HPV in oral cavity—is a possible approach, but it would be more subjective and potentially misleading, since the strength of the published evidence is controversial. The high threshold of evidence we used might have prevented us from addressing cancers for which evidence of an infectious link is rapidly emerging. Nevertheless, we aimed to not exaggerate the importance of infections in cancer. Estimates of relative risks and infection prevalence were always derived from the same review, or from systematic or comprehensive reviews that we updated when necessary. Cancer incidence data were derived from GLOBOCAN 2008, or calculated using a consistent method when GLOBOCAN estimates were not available.

For most PAF calculations, we used estimates of infection prevalence based on case series rather than

	Number of new cases in 2008	Number attributable to infection	PAF (%)	Number attributable to infection, by sex		Number attributable to infection, by age group			Number attributable to infection, by development status	
				Male	Female	<50 years	50–69 years	≥70 years	Less developed regions	More developed regions
Carcinoma										
Non-cardia gastric	870 000	650 000	74·7%	410 000	240 000	82 000	290 000	270 000	470 000	180 000
Liver†	750 000	580 000	76·9%	400 000	170 000	130 000	280 000	180 000	510 000	69 000
Cervix uteri	530 000	530 000	100·0%	0	530 000	250 000	220 000	59 000	450 000	77 000
Vulva	27 000	12 000	43·0%	0	12 000	1700	3900	6000	4100	7500
Anus	27 000	24 000	88·0%	11 000	13 000	5100	10 000	8300	12 000	12 000
Penis	22 000	11 000	50·0%	11 000	0	2500	4800	3500	7600	3200
Vagina	13 000	9000	70·0%	0	9000	2000	4000	3100	5700	3400
Oropharynx	85 000	22 000	25·6%	17 000	4400	4300	13 000	4600	6400	15 000
Nasopharynx	84 000	72 000	85·5%	49 000	23 000	31 000	32 000	9200	66 000	5900
Bladder	260 000	6000	2·3%	4600	1400	1200	3400	1400	6000	0
Lymphoma and leukaemia										
Hodgkin's	68 000	33 000	49·1%	20 000	13 000	23 000	6700	3500	23 000	10 000
Non-Hodgkin gastric	18 000	13 000	74·1%	7400	5800	3900	5000	4400	6500	6700
Burkitt	11 000	6800	62·5%	4000	2800	6300	290	210	6300	530
HCV-associated non-Hodgkin	360 000	29 000	8·2%	17 000	13 000	9500	11 000	8800	18 000	11 000
Adult T-cell	2100	2100	100·0%	1200	900	580	980	580	660	1500
Kaposi's sarcoma	43 000	43 000	100·0%	29 000	14 000	30 000	7600	4700	39 000	4100
Total infectious-disease-related sites	3 200 000	2 000 000	64·4%	990 000	1 100 000	580 000	890 000	560 000	1 600 000	410 000

PAF=population attributable fraction. HCV=hepatitis C virus. *Numbers are rounded to two significant digits. †Including cholangiocarcinoma.

Table 4: **Number of new cancer cases* in 2008 attributable to infectious agents, by anatomic site**

prevalence in the general population. This choice was made for two reasons. First, there are few large, high-quality, population-based surveys of infection prevalence that are representative of the general population. Such surveys tend to oversample young people or those at low risk for specific infections. Undertaking high-quality population-based surveys is a long and difficult process and has not been done at all in many countries, particularly in less developed regions. Cancer case series, which are available in less developed countries, are usually representative of the population served by the hospital. Patients with cancer need expert care, and the severity of the disease makes it very likely that these patients will seek appropriate treatment in specialised centres. Second, population-based surveys often use less sensitive or specific measurement methods than case series, because the best testing methods are often expensive, invasive, and less feasible on a large scale. For example, causation is difficult to determine from serology in cases of EBV-related or HPV-related cancers; more than 90% of the population is positive for EBV, and HPV serology is not site-specific. However, patients included in case series undergo many tests, often including direct detection of infectious agents and even gene expression in tumour tissue. Therefore, we are confident that calculating prevalence from case series increases the validity of PAF estimates.

The need to avoid the effect of time trends in infection prevalence when selecting the case series on which we based our calculations was also considered, but did not seem to be an issue, at least for the four main infections that drive most of the global PAF. Although improved living conditions have led to a steadily decreasing prevalence of *H pylori* infection in many populations, prevalence in gastric cancer cases from nested case–control studies or other epidemiological studies seems to be very stable, around 90%, with no detectable secular trends. The case series we selected for estimating the prevalence of hepatocellular carcinoma attributable to HBV and HCV used only second-generation or third-generation ELISA for detection of HCV, with a total of 37 000 cases from 132 studies published from 1992 to 2009. The low sensitivity of first-generation anti-HCV ELISA has long been recognised and seems to differ between cases and controls.[13] Most studies of HPV-related cancers have been done in the past 15 years using DNA detection techniques, and relevant studies of oropharyngeal cancers are even more recent.

Our approach has several limitations. First, some uncertainty in cancer incidence and infection prevalence is inherent in our estimations of PAF. Our attempt to obtain global estimates of infection prevalence by pooling local data sources inevitably requires extrapolation to countries with sparse data on cancer incidence or

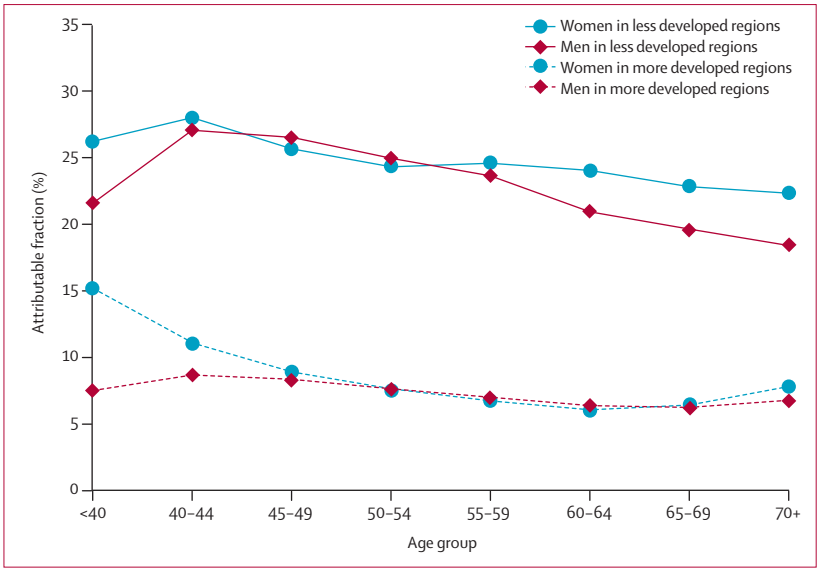

Figure 3: **Relative percentage of new cancer cases in 2008 attributable to infection, by sex, age group, and development status**

Panel 2: **Research in context**

Systematic review

The relation between infectious agents and cancer was the subject of a comprehensive literature review done as part of the IARC Monographs programme. Global cancer incidence and mortality data were synthesised by the GLOBOCAN project to provide estimates of the global burden of cancer. We used these data sources to estimate the global burden of cancer due to infection, relying on existing systematic reviews of the literature to provide the quantitative inputs (relative risk and infection prevalence) required for calculation of attributable fractions. Previous global estimates of the proportion of cancers attributable to infection were done for 1990[4] and for 2002.[5] Country-specific estimates have been provided for the UK,[18] South Korea,[19] and China.[20]

Interpretation

The present review extends previous findings by showing wide geographical variation in the fraction of cancers attributable to infection. It also underscores the importance of HPV, *Helicobacter pylori*, HBV, and HCV as cancer-related infectious agents. Since infections are an important and preventable cause of cancer worldwide, clinicians should support the implementation of available strategies for prevention—ie, vaccination against HBV and HPV, use of safe injection practices, and avoidance of parenteral treatment when oral treatment is available. Clinicians should also closely follow and, if possible, contribute to progress in areas where early detection of infection (eg, HPV) or treatment (eg, HCV and *H pylori*) could diminish cancer sequelae. Public health doctors and cancer-control specialists should appreciate the importance of infectious causes of cancer in different regions and age groups, particularly in low-income and middle-income populations. The 2011 UN high-level meeting on non-communicable diseases highlighted the growing global agenda for prevention and control of non-communicable diseases. Although cancer is considered a major non-communicable disease, a sizable proportion of its causation is infectious and simple non-communicable disease paradigms will not be sufficient.

risk-factor prevalence. The main risk of this extrapolation is that estimates derived from a small amount of data are applied to a larger population that is substantially different, with possible amplification of bias. The weighting scheme we used to estimate regional infection prevalence might give undue weight to larger surveys done in smaller countries.

A second limitation is that strong assumptions were required for the calculations. For example, we assumed that relative risks for infection were constant across populations and sexes—a common assumption in epidemiology. Generally, this assumption is not true when comparing populations or sexes with widely different baseline cancer incidence rates, since the multiplicative assumption behind relative risk estimation is only an approximation. Our PAF calculations were not strongly dependent on the constant relative risk assumption because we used the retrospective formula based on cancer cases, and the relative risks were uniformly large. A change in relative risk from 10 to 20, for example, makes only a 5% difference in the PAF estimate. For *H pylori*, HPV, HBV, and HCV, the order of magnitude of relative risk is generally constant worldwide when other known risk factors have been controlled for, so the assumption of a constant relative risk should not lead to substantial error.

A third limitation is the lack of high-quality epidemiological data for some of the cancer sites in this study (eg, EBV-related cancers, such as nasopharyngeal carcinoma), in areas of low cancer incidence. We based our estimates on the most recent and least subjective evidence, but the lack of data inevitably leads to uncertainty in the estimates.

Some of the assumptions used in our calculations were conservative. We restricted the effect of HPV in head and neck cancer to the oropharynx and base of the tongue, where the epidemiological and mechanistic evidence for a causal effect is strongest. HPV might be associated with other head and neck cancers, but this is impossible to quantify with current data. For *H pylori*, we based PAF calculations on a relative risk of 5·9; although this is the best estimate available, evidence from prospective studies and studies using western blot rather than ELISA suggests that it might be higher. Such studies yielded relative risks greater than 10,[21–25] which would increase the proportion of non-cardia gastric cancer attributable to *H pylori* from 75% to 90%. Likewise, we estimated that 75% of diffuse large B-cell lymphoma of gastric location is due to *H pylori*, but the attributable fraction could be nearer to 100%. The discovery of new associations between infections, particularly viruses, and cancer has been anticipated; however, studies have either disclosed associations with very rare cancers (eg, Merkel-cell carcinoma) or are yet to provide conclusive results (eg, for cutaneous HPV types and non-melanomatous skin cancer). Nevertheless, undiscovered associations could exist, which is another reason to conclude that our results probably underestimate the true burden of infection-associated cancers.

Attributable-risk calculations can be done for any environmental exposure, but are most useful, from a

public health perspective, when relative risks are large and when interventions to reduce population exposure are feasible. Many infection-related cancers are preventable (panel 2), particularly those associated with HPV, *H pylori*, HBV, and HCV. Prophylactic vaccines have shown nearly 100% efficacy in preventing precancerous lesions of the cervix due to HPV types 16 and 18, among previously uninfected individuals. In Taiwan, the incidence of hepatocellular carcinoma in children and adolescents has been substantially reduced by a combination of immunoglobulin given at birth to prevent vertical transmission from mother to child at birth and childhood HBV vaccination.[1] The current WHO recommendation is to vaccinate all infants against HBV as soon as possible after birth.[26] Although no vaccine is available for HCV, iatrogenic transmission can be avoided with safer practices for injection and blood transfusion, and preference for oral drug delivery over injections where available. *H pylori* is a treatable infection, although the feasibility, effectiveness, and safety of large-scale eradication of *H pylori* infection in different age groups is not yet clear. Our finding that *H pylori* accounts for 46% of infection-associated cancers in more developed areas might reflect lower investment in research on prevention of gastric cancer compared with cervical and liver cancers. Such considerations underscore the difference between what is theoretically preventable, according to the assumptions of the PAF calculations, and what is preventable in practice. The importance of time must also be acknowledged; preventing infection-associated cancers in 2008 would have required intervention programmes many decades earlier.

In view of the high mortality rate of infection-associated cancers, the fraction of cancer deaths attributable to infections is probably higher than the 16·1% that our study generated. Although a full investigation of cancer death due to infection is beyond the scope of this report, we can estimate the mortality burden by applying the PAFs to the 7·5 million cancer deaths that occurred in 2008. These calculations suggest that 1·5 million cancer deaths were attributable to infectious agents, or roughly one in five deaths due to cancer worldwide.

Contributors

SF, DF, MP, and CDM conceived and designed the study. JF provided cancer incidence estimates adapted from the GLOBOCAN 2008 database. MP, JV, and FB contributed to data collection and data analysis. CDM and MP wrote the manuscript. All authors contributed to the interpretation of data and approved the final manuscript.

Conflicts of interest

We declare that we have no conflicts of interest.

Acknowledgments

CDM was funded by Fondation Innovations en Infectiologie (FINOVI) and the Bill & Melinda Gates Foundation (BMGF), USA (grant number 35537). We thank Veronique Chabanis for technical assistance.

References

1 Chang MH, You SL, Chen CJ, et al. Decreased incidence of hepatocellular carcinoma in hepatitis B vaccines: a 20-year follow-up study. *J Natl Cancer Inst* 2009; **101:** 1348–55.
2 IARC. Monographs on the evaluation of carcinogenic risks to humans, volume 100. A review of carcinogen—Part B: biological agents. Lyon: International Agency for Research on Cancer, 2011.
3 Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010; **127:** 2893–917.
4 Pisani P, Parkin DM, Muñoz N, Ferlay J. Cancer and infection: estimates of the attributable fraction in 1990. *Cancer Epidemiol Biomarkers Prev* 1997; **6:** 387–400.
5 Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* 2006; **118:** 3030–44.
6 Bouvard V, Baan R, Straif K, et al. A review of human carcinogens–Part B: biological agents. *Lancet Oncol* 2009; **10:** 321–22.
7 UN. World population prospects: the 2008 revision highlights. New York; United Nations, 2009.
8 Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. GLOBOCAN 2008 v1.2, cancer incidence and mortality worldwide: IARC CancerBase 10. Lyon: International Agency for Research on Cancer, 2010.
9 Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB. Cancer incidence in five continents, vol VIII. IARC scientific publications 155. Lyon: International Agency for Research on Cancer, 2002.
10 Curado MP, Edwards B, Shin HR, et al. Cancer incidence in five continents, vol IX. IARC scientific publications 160. Lyon: International Agency for Research on Cancer, 2007.
11 Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am J Epidemiol* 1974; **99:** 325–32.
12 Helicobacter and Cancer Collaborative Group. Gastric cancer and *Helicobacter pylori*: a combined analysis of 12 case control studies nested within prospective cohorts. *Gut* 2001; **49:** 347–53.
13 Donato F, Boffetta P, Puoti M. A meta-analysis of epidemiological studies on the combined effect of hepatitis B and C virus infections in causing hepatocellular carcinoma. *Int J Cancer* 1998; **75:** 347–54.
14 Dondog B, Lise M, Dondov O, Baldandorj B, Franceschi S. Hepatitis B and C virus infections in hepatocellular carcinoma and cirrhosis in Mongolia. *Eur J Cancer Prev* 2011; **20:** 33–39.
15 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:** 2078–85.
16 Backes DM, Kurman RJ, Pimenta JM, Smith JS. Systematic review of human papillomavirus prevalence in invasive penile cancer. *Cancer Causes Control* 2009; **20:** 449–57.
17 De Vuyst H, Clifford GM, Nascimento MC, Madeleine MM, Franceschi S. 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:** 1626–36.
18 Parkin DM. Cancers attributable to infections in the UK in 2010. *Br J Cancer* 2011; **105** (suppl 2)**:** 49–56.
19 Shin A, Park S, Shin HR, et al. Population attributable fraction of infection-related cancers in Korea. *Ann Oncol* 2011; **22:** 1435–42.
20 Xiang W, Shi JF, Li P, et al. Estimation of cancer cases and deaths attributable to infection in China. *Cancer Causes Control* 2011; **22:** 1153–61.
21 Mitchell H, English DR, Elliott F, 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:** 903–10.
22 Ekstrom AM, Held M, Hansson LE, Engstrand L, Nyren O. *Helicobacter pylori* in gastric cancer established by CagA immunoblot as a marker of past infection. *Gastroenterology* 2001; **121:** 784–91.
23 Brenner H, Arndt V, Stegmaier C, Ziegler H, Rothenbacher D. Is *Helicobacter pylori* infection a necessary condition for noncardia gastric cancer? *Am J Epidemiol* 2004; **159:** 252–58.
24 Herrera V, Parsonnet J. *Helicobacter pylori* and gastric adenocarcinoma. *Clin Microbiol Infect* 2009; **15:** 971–76.
25 Siman JH, Engstrand L, Berglund G, Forsgren A, Floren CH. *Helicobacter pylori* and CagA seropositivity and its association with gastric and oesophageal carcinoma. *Scand J Gastroenterol* 2007; **42:** 933–40.
26 WHO. Weekly epidemiological report: hepatitis B vaccines. Geneva: World Health Organization, 2010.