Geospatial Health 2016; volume 11:396

pagepress

How much incident lung cancer was missed globally in 2012? An ecological country-level study

Benn Sartorius,[1] Kurt Sartorius[1,2]

[1]School of Nursing and Public Health, University of KwaZulu-Natal, Durban;
[2]Faculty of Commerce, University of the Witwatersrand, Johannesburg, South Africa

Abstract

Lung cancer incidence is increasing in many low-to-middle-income countries and is significantly under-reported in Africa, which could potentially mislead policy makers when prioritising disease burden. We employed an ecological correlation study design using country-level lung cancer incidence data and associated determinant data. Lagged prevalence of smoking and other exposure data were used to account for exposure-disease latency. A multivariable Poisson model was employed to estimate missed lung cancer in countries lacking incidence data. Projections were further refined to remove potential deaths from infectious/external competing causes. Global lung cancer incidence was much lower among females *vs* males (13.6 *vs* 34.2 per 100,000). Distinct spatial heterogeneity was observed for incident lung cancer and appeared concentrated in contiguous regions. Our model predicted a revised global lung cancer incidence in 2012 of 23.6 compared to the Globocan 2012 estimate of 23.1, amounting to ~38,101 missed cases (95% confidence interval: 28,489-47,713). The largest relative under-estimation was predicted for Africa, Central America and the Indian Ocean regions (Comoros, Madagascar, Mauritius, Mayotte, Reunion, Seychelles). Our results suggest substantial under-reporting of lung cancer incidence, specifically in developing countries (*e.g.* Africa). The *missed* cost of treating these cases could amount to >US$ 130 million based on recent developing setting costs for treating earlier stage lung cancer. The full cost is not only under-estimated, but also requires substantial additional social/family inputs as evidenced in more developed settings like the European Union. This represents a major public health problem in developing settings (*e.g.* Africa) with limited healthcare resources.

Correspondence: Benn Sartorius, Discipline of Public Health Medicine, 2nd floor George Campbell Building, Howard College Campus, University of KwaZulu-Natal, Durban 4001, South Africa.
Tel: +27.031.2604459 - Fax: +27.031.2604211.
E-mail: Sartorius@ukzn.ac.za

Key words: Under reporting; Incident lung cancer; Ecological determinant model.

Received for publication: 3 July 2015.
Revision received: 12 October 2015.
Accepted for publication: 19 October 2015.

Geospatial Health 2016; 11:396
doi:10.4081/gh.2016.396

Introduction

Lung cancer remains the most common cancer worldwide, and recent surveys illustrate persistent high levels of lung cancer and an increasing trends of female prevalence in many developed regions like North America and Europe (WHO-IARC, 2014; Islami *et al*., 2015; Lortet-Tieulent *et al*., 2014). Many low to middle income countries also show increasing levels of male-led incidence in regions like Eastern and Central Europe, particularly in Hungary, Poland and Serbia (WHO-IARC, 2014; Lortet-Tieulent *et al*., 2014; Tesfaye *et al*., 2007). A recent study in 27 countries belonging to the European Union (EU) indicates that the economic costs of lung cancer for 2009 were €18.8 bn and the biggest contributor in economic productivity (Edge *et al*., 2010). Meanwhile, the potential cost of lung cancer in developing countries is projected to increase further because of the trend of rising incidence in the next decade (Farnsworth *et al*., 2004; Murray, 1997; Nishida and Kudo, 2013). Recent estimates report that the developing countries account for 58 per cent of the global burden of lung cancer incidence (WHO-IARC, 2014) and contain an estimated 1.38 billion smokers, who are associated with a 7-fold higher risk of acquiring lung cancer (Nishida and Kudo, 2013). Given the high economic cost, rising lung cancer incidence in developing countries is likely to further impact available health resources due to the limited resources of these regions (Shorrocks, 2013).

The complexity of lung cancer incidence is illustrated by the stratified nature of its biology and geno-environmental determinants (Minagawa *et al*., 2007; Tesfaye *et al*., 2007) that show both inter- and intra-country variations. The Globocan 2012 report shows marked inter-country incidence differences that are compounded by intra-country differences across regions, ethnic groups (WHO-IARC, 2014; Lortet-Tieulent *et al*., 2014), gender (Islami *et al*., 2015), a long latency period (30 years) and the degree of inhalation (Farnsworth *et al*., 2004). Although smoking is widely accepted as the primary agent that triggers lung cancer (Bruce *et al*., 2000; Cohen *et al*., 2005), evidence exists that shows that the disease is also linked to other social determinants like passive smoking, indoor cooking, occupational hazards, motor vehicle emissions and industrial pollution (Lary *et al*., 2014; Youlden *et al*., 2008). A combination of factors therefore suggests diverse inter- and intra-country differences in the incidence of lung cancer that are further magnified by the inequality gap between socioeconomic groups (Deaton, 2013). In addition, the management of lung cancer is compromised because its incidence may be significantly under-reported in (many) developing countries because of lacking data or poor quality thereof (WHO-IARC, 2014; Ferlay *et al*., 2010). Although the geography of lung cancer incidence is well recorded in the literature (Tesfaye *et al*., 2007), the under-reporting of its incidence in many countries could potentially mislead policy makers when prioritising disease burden. A need therefore exists to identify geographic areas that lack or significantly under-report lung cancer as this has implications for disease prioritisation, resource allocation and tailored policy interventions (Islami *et al*., 2015; Tesfaye *et al*., 2007). Lung cancer management in developing countries will continue to be compromised by its complexity as well as under-estimation of the impact of the disease because of poor or lacking data. The aim of this study was to quantify potential under-estimation of lung cancer using Globocan 2012 data and available lagged determinant (risk factor) data. The level and implications of this under-estimation are discussed in order to propose some relevant policy implications.

Materials and Methods

Incidence data

Age- and gender-standardised lung cancer incidence as well as raw counts for 184 countries were extracted from the GLOBOCAN 2012 database (WHO-IARC, 2014). A detailed description of the data sources, methodologies and potential limitations have been published previously (WHO-IARC, 2014). Data from countries are classified according to availability and quality, namely: A [high quality national data or high quality regional (coverage greater than 50%)], B [high quality^2 regional (coverage between 10% and 50%)], C [high quality^2 regional (coverage lower than 10%)], D [national data (rates)], E [regional data (rates)], F (frequency data), G (no data).

Determinants

The following key determinants were included in the analysis. Published data for historical levels and trends in age standardised prevalence of smoking by country were utilised (Ng *et al*., 2014). These data are for 187 countries and estimated age-standardised prevalence of smoking in 1980, 1996, 2006 and 2012. We used estimates from 1980 in our model as this would most closely correspond with the 30-year latency period described below. Missing smoking prevalence estimates for countries were imputed using a Bayesian spatial conditional autoregressive (CAR) model to use neighbouring country values. Historical outdoor air pollution data with reference to particulate matter concentrations less than 10 microns in diameter (PM_{10}) expressed as micrograms per m^3 were extracted from the World Bank Data repository for the period 1990-2012 (World Bank, 2013). We used estimates for 209 countries from 1990 in our model (data were not available prior to that in the repository). Proportions of households using solid fuel for cooking (indoor pollution) by country in 1990, 2000 and 2010 were extracted from Bonjour *et al*. (2013) and these are based on data extracted from the World Health Organization (WHO) Household Energy Database (WHO, 2009, 2012). Data used by Bonjour *et al*. (2013) covered 155 countries with at least one survey per country between 1974 and 2010. For countries lacking solid-fuel data but classified as high-income countries, they assumed the proportion using solid fuels for cooking to be <5% (Bonjour *et al*., 2013). We used estimates for 184 countries from 1990 in our model. Gross domestic product (GDP) data per capita (in US$) by country in 1980 were also extracted from the World Bank Data repository (World Bank, 2013). Furthermore, mortality from competing causes of death (specifically communicable disease and injury) were extracted for each country from the Global Burden of Disease Project Study 2013 (Naghavi *et al*., 2015) to realistically reduce projected under-estimation from the model as some incident lung cancer would not have been realised (Figure 1).

Latency period between exposure and lung cancer

A review of available literature suggests an approximate 30-year population latency period between smoking prevalence and subsequent lung cancer mortality (Burbank, 1972; Devesa *et al*., 1987; Finkelstein, 1991; Polednak, 1974; Weiss, 1997). This estimate was used to estimate the period to be used between historical population level smoking exposure and its subsequent impact on current lung cancer incidence. This study employed an ecological country-level correlation design.

Data analysis

Spatial analysis

Local indicators of spatial autocorrelation (LISA) (Anselin, 1996) were estimated using the freely available GeoDa software (Anselin *et al*., 2006). Local Moran's I statistic based on rates (cases and associated populations at risk) was used to identify the existence of significant spatial clustering of lung cancer by gender. GeoDa implements the recommended Empirical Bayes (EB) standardisation procedure (Assuncao and Reis, 1999) in its global Moran scatter plot and LISA maps *i.e*. standardisation of the raw rates. Significance was set at 5% after 99,999 iterations.

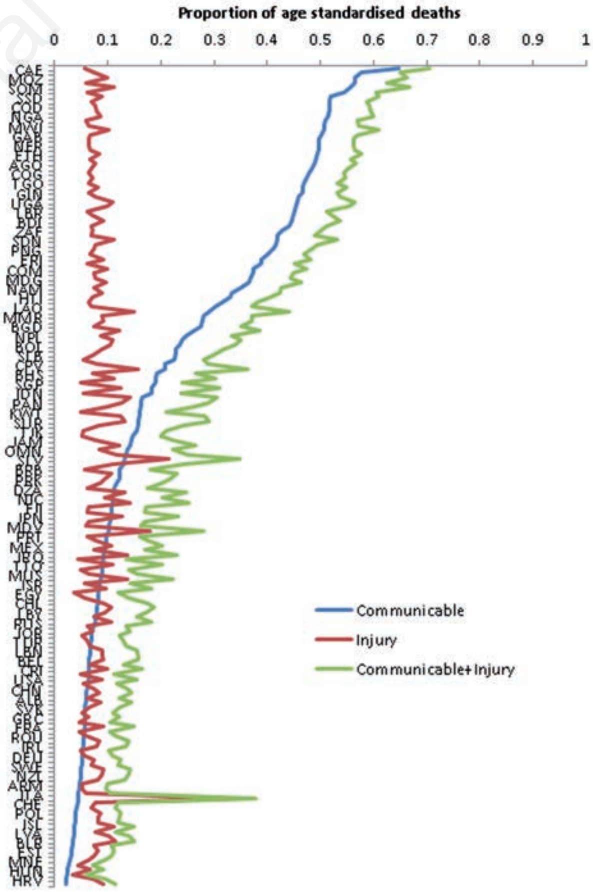

Figure 1. Fraction of mortality due to competing infectious disease and injury by country. The values on the vertical axis represent ISO 3166 country codes.

Statistical analyses

Analyses were performed using STATA software version 13.0 (StataCorp, 2011). Using observed numbers of age standard incident lung cancer cases and associated population at risk, we calculated incidence rates and associated 95% Poisson confidence limits. Country level lung cancer incidence was considered significantly above average if the lower 95% confidence interval (CI) limit (α=0.025) of the incidence proportion would be above the global average (Pickle *et al.*, 1987). We employed a regression framework to quantify and identify determinant associated with lung cancer utilised previously. We employed an ecological (country-level) generalised linear Poisson modelling approach with robust standard errors (to correctly adjust the standard errors and not over-estimate significance) to estimate risk ratios (RR) for each determinant *vs* lung cancer incidence. Factors significant at the 10% level, based on the bivariate regressions, were selected for inclusion into a multivariable model. The final multivariable model was then used to predict *potentially missed* lung cancer cases based on observed covariate values, along with 95% confidence intervals (uncertainty). We only predicted missing lung cancer incidence in countries with potentially under-reported or missing national level cancer data as described in the outcome data section above.

Results

Global incidence

The overall global estimated incidence rate for lung cancer in 2012 was found to be 23.1 per 100,000 population (before accounting for potential

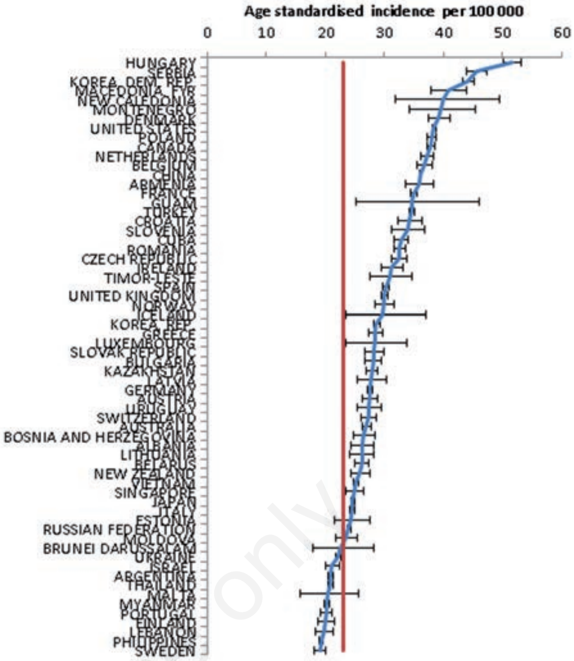

Figure 2. Countries above the global average age standardised lung cancer incidence (plus 95% confidence intervals) per 100,000 population in 2012. The horizontal red line represents the overall global incidence.

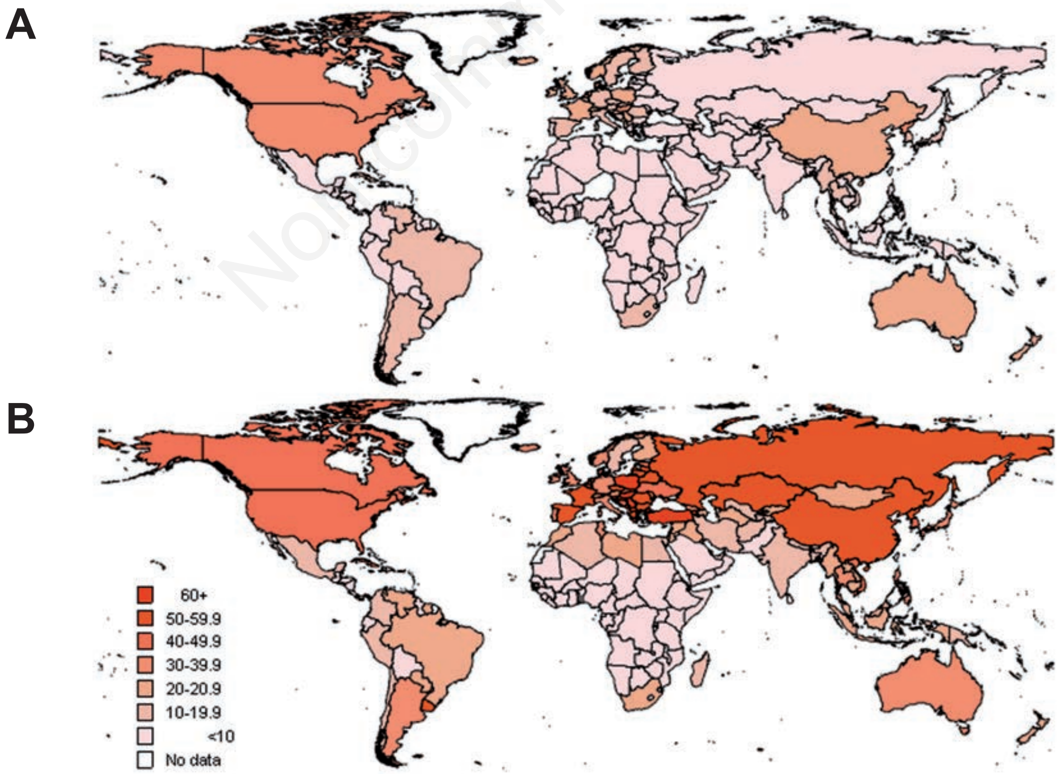

Figure 3. Map depicting observed age-standardised lung cancer incidence per 100,000 population by gender (A, female; B, male) in 2012.

under-reporting) with male incidence more than twice that of females. The highest overall age-standardised incidence of lung cancer was in Hungary (51.6 per 100,000) followed by Serbia, Democratic Republic of Korea and Macedonia with rates of 45.6, 44.2 and 40.8 per 100,000 respectively (Figure 2). Six of the top 20 are classified as developed countries. China and USA were the leading countries in terms of absolute burden of new cases in 2012 with 652,842 and 214,226 cases, respectively, followed by Japan (94,855), India (70,275) and Russia (55,805). Very low incidence rates were observed in Middle and Western Africa (likely as a result of under-estimation due to poor data quality), with the lowest incidences estimated in Niger (0.2 per 100,000), Tanzania (0.7), Malawi (0.9) and the Democratic Republic of Congo (1.0).

Spatial risk by gender

Global and country-level incidence was generally much lower among females compared to males (globally 13.6 *vs* 34.2 per 100,000, respectively). The highest female incidence rates were mainly contained in developed countries like North America [Canada (rank 2) and USA (3)] and Western Europe [Denmark (1), Hungary (5), Netherlands (6), Iceland (7), Ireland (8), Norway (9) and United Kingdom (10)] (Figure 3A). The highest burden of absolute incident female cases was observed in China (193,347) and USA (102,172).

Conversely, high levels of male incidence (Figure 3B) were more widespread and included Europe, North America, the Eastern Mediterranean and Asia while lower levels were reflected in Africa (except for South Africa). The highest male incidence was largely concentrated in south-eastern and south-western Europe and Central Asia. The 10 leading countries (in descending order) were: Hungary, Armenia, Macedonia, Serbia, Turkey, Montenegro, Poland, Kazakhstan, Romania and the Democratic Republic of Korea. The highest burden of absolute incident male cases was again observed in China (459,495) and USA (112,054).

Prominent differences in the clustering of gender-specific incidence (based on neighbouring contiguity) were also observed (Figure 4). Significant spatial clustering of higher male lung cancer incidence was largely concentrated in Central-Eastern Europe and Northern Asia (Russia) with isolated significant high-risk countries in Africa (South Africa, Libya, and Morocco) and South America (Argentina). Significant spatial clustering of higher female incidence was largely concentrated in Central-Eastern Europe with isolated significant high-risk countries in Africa (South Africa, Morocco, *i.e.* similar to males) and Cuba. Significant clustering of low lung cancer incidence was largely confined to Africa for both genders (again see potential under-estimation in data and methods section).

Lung cancer incidence and associated historical determinants

Summary statistics for the selected determinants are presented in Table 1. Global prevalence of age standardised smoking in 1980 was estimated to be 25.9% [95% CI: 25.1, 26.6] by Ng and colleagues (2014). However, this varies by continent, with the highest historical prevalence of smoking and historical GDP per capita in Europe (~31% and US$ 12,389, respectively). The highest historical exposure to indoor pollution was observed in Africa (~77%). A detailed summary of all the determinants is shown in Table 1.

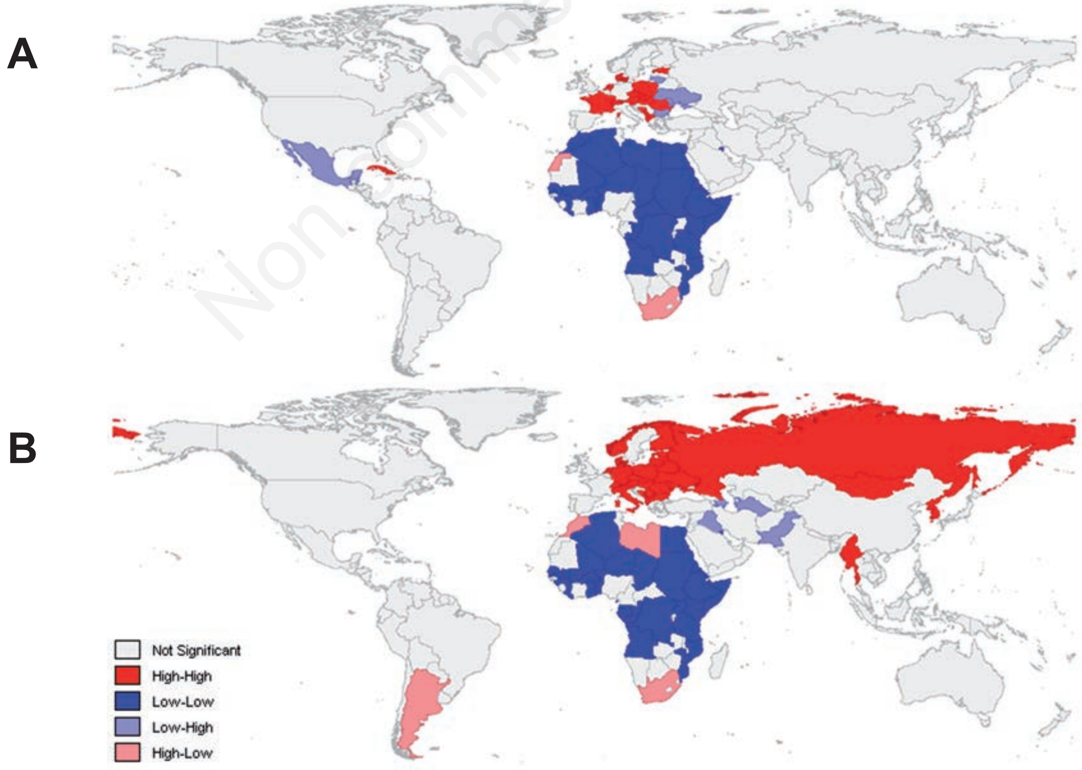

Figure 4. Significant (P<0.05) spatial clustering of age-standardised lung cancer incidence (per 100,000 population) by gender (A, female; B, male) in 2012.

Causal associations

Age-standardised smoking prevalence in 1980 as well as indoor and outdoor pollution levels in 1990 were all significantly associated with lung cancer incidence based on bivariate regressions. Figure 5 shows the strong positive correlation between lagged prevalence of smoking (in 1980) and observed lung cancer incidence in 2012.

Following multivariable adjustment, increasing smoking prevalence and outdoor pollution remained significant risk factors for current age-standardised lung cancer incidence with relative risks (RR) of 1.12 and 1.01, respectively (Table 2) for each unit increase respectively. Figure 6 below depicts the distribution of predicted age standardised incidence of lung cancer for countries with lack of, or poor (under-estimated) data.

Under-estimation of lung cancer incidence by region, country and additional direct cost

Predictions from our model suggest that the true global incidence of

Table 1. Summary of historical determinant values by continent.

Continent	Age-standardised smoking prevalence (%) in 1980	Indoor pollution (% of households using solid fuel for cooking) in 1990	Outdoor pollution (PM$_{10}$ per m^3) in 1990	GDP per capita (US$) in 1980
Africa	12.74 (11.23,14.26)	77.08 (68.88,85.27)	60.88 (48.06,73.71)	821 (517,1126)
Americas*	17.11 (13.93,20.3)	32.83 (24.91,40.74)	42.93 (35.35,50.51)	3171 (2100,4242)
Asia	23.66 (21.28,26.03)	43.2 (32.32,54.09)	84.81 (66.19,103.44)	6571 (2508,10,635)
Europe	30.78 (29.08,32.48)	13.38 (8.16,18.59)	51.21 (43.73,58.69)	12,389 (8267,16,511)
Oceania°	30.33 (25.01,35.64)	44.67 (15.9,73.44)	31.24 (23.85,38.62)	4028 (1097,6958)

GDP, gross domestic product. Values are expressed as mean (95% confidence interval). *Includes North, Central and South America; °comprising Australia and proximate islands.

Table 2. Bivariate and multivariable adjusted (using a Poisson regression model) determinants associated with lung cancer in 2012.

Determinant	Unadjusted risk ratio (95% CI)	P	Adjusted risk ratio (95% CI)	P
Age-standardised prevalence of smoking in 1980 (%)	1.09 (1.03,1.16)	0.004	1.12 (1.05,1.18)	<0.001
Outdoor pollution in 1990 (PM$_{10}$ per m^3)	1.02 (1.01,1.03)	0.001	1.01 (1.00,1.02)	0.020
Indoor pollution (% of households using solid fuel for cooking)	1.03 (1.01,1.04)	<0.001	1.01 (0.99,1.00)	0.193
GDP per capita (US$)	0.999 (0.999,0.999)	0.007	0.999 (0.999,1.000)	0.613

CI, confidence interval; GDP, gross domestic product.

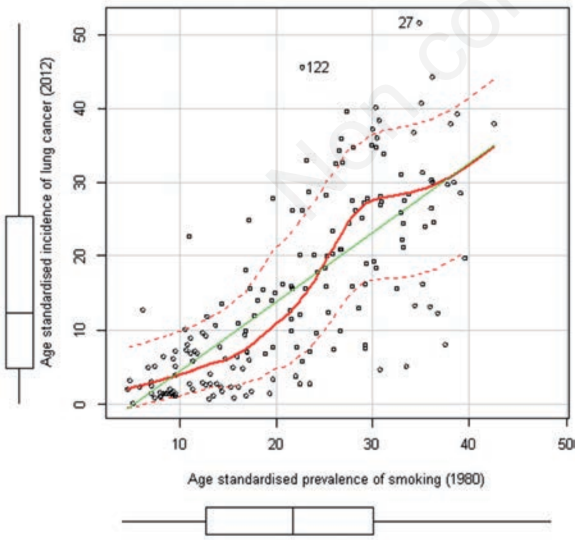

Figure 5. Relationship between lagged age-standardised prevalence of smoking (1980) and current observed age-standardised lung cancer incidence (2012). Solid red line represents locally weighted scatterplot smoothing [LOWESS] fitted line, solid green line stands for linear trend, while dashed red line is the 95% confidence limits around LOWESS fitted line. Marginal boxplots are also presented alongside respective axis.

Table 3. Top ten largest absolute number of potentially missed incident lung cancer cases by country in 2012.

Country	Region	Missed cases*	95% CI
Nepal	South Asia	2902	2833-2970
Bolivia	South America	1463	1401-1526
Papua New Guinea	Pacific	803	763-844
Greece	South East Europe	587	569-604
Congo	Central Africa	526	497-554
Lao	South East Asia	439	417-461
Kosovo	South East Europe	380	342-420
Burundi	Central Africa	225	204-247
Sierra Leone	Western Africa	147	129-166
Macao, China	East Asia	118	97-141

CI, confidence interval. *Absolute number.

lung cancer in 2012 may have been 23.6 per 100,000 (95% CI: 23.4-23.7) even after adjustment for competing causes of mortality compared to reported estimates of 23.1 per 100,000 *i.e.* ~2.2% or 0.5 additional incident cases per 100,000 population globally. In absolute terms, this could have resulted in potentially 38 101 missed cases (95% CI: 28 489-47 713) in 2012.

The under-estimation by WHO region, illustrated in Figure 7A, compares the ratio of predicted incidence by the model *vs* observed incidence where a value significantly in excess of 1 suggests under-estimation. Our model suggests that lung cancer incidence has been significantly under-estimated in a number of regions, particularly Central Africa, Central America, Western Africa, Eastern Africa and the Pacific region. In absolute numbers, the burden of under-estimated cases is largest in South Asia (5900 cases) followed by Central America (5812 cases) and South America (3312 cases).

The under-estimation of the burden of lung cancer incidence, illustrated in Figure 7B, compares the ranked predicted incidence to the observed incidence in all the WHO sub-regions. The largest absolute burden increase was predicted in Central America with potentially 3.5

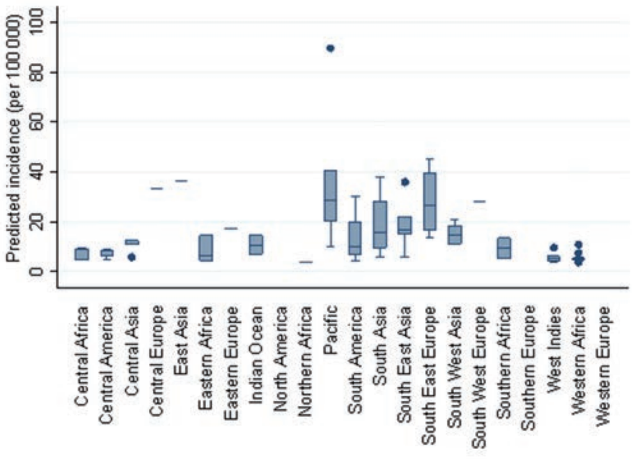

Figure 6. Model predicted age-standardised incidence of lung cancer in countries with missing or poor quality data in 2012.

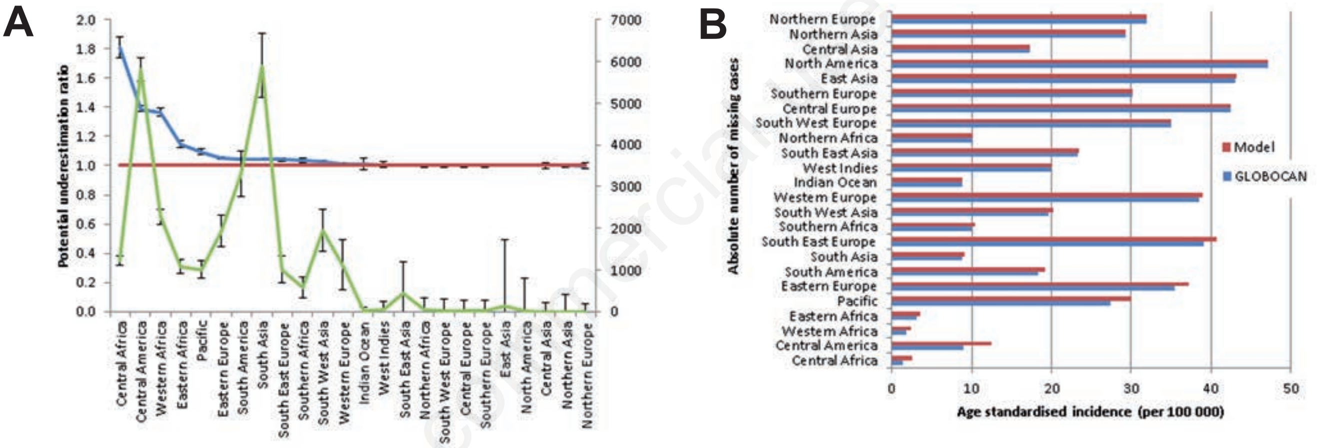

Figure 7. A) Ratio of and comparative absolute age-standardised incidence predicted by the model *vs* observed age-standardised incidence by region in 2012. Blue line represents the relative underestimation ratio (primary x axis); red line is the null underestimation, *i.e.* ratio=1 (primary y axis); green line shows the absolute number of missing cases (secondary y axis). B) Under-estimation of the burden of lung cancer incidence.

Table 4. Top ten largest relative number of potentially missed incident lung cancer cases by country in 2012.

Country	Region	Number of cases*	Predicted number of cases°	Relative under-estimation	95% CI
Cabo Verde	Western Africa	5	16.1	3.3	1.9-5.4
Bolivia	South America	659	2121.2	3.2	3.1-3.4
Comoros	Indian Ocean	11	26.9	2.5	1.7-3.7
Sierra Leone	Western Africa	110	257.0	2.3	2.1-2.6
Papua New Guinea	Pacific	706	1508.4	2.1	2.0-2.2
Burundi	Central Africa	206	431.0	2.1	1.9-2.3
Djibouti	Eastern Africa	29	58.8	2.1	1.6-2.7
Nepal	South Asia	4160	7060.6	1.7	1.7-1.7
Solomon islands	Pacific	51	85.9	1.7	1.3-2.1
Congo	Central Africa	809	1334.7	1.7	1.6-1.7

CI, confidence interval. *Observed number of age standardised cases; °predicted by the model.

missed lung cancer cases per 100,000 populations. This was followed by the Pacific Region (2.6 missed cases per 100,000), eastern Europe (1.9 missed cases per 100,000), southeastern Europe (1.6 missed cases per 100,000) and central Africa (1.1 missed cases per 100,000). The model suggests that the highest, absolute numbers of potentially missed cases of lung cancer were in Nepal, Bolivia and Papua New Guinea (Table 3). In relative terms, the highest relative under-estimation was predicted for Cabo Verde followed by Bolivia, Comoros and Sierra Leone (Table 4).

Discussion

Our results confirm some common established determinants of lung cancer that include lagged smoking prevalence (Allison, 1978), indoor (Bruce *et al*., 2000) and outdoor pollution (Cohen *et al*., 2005) in order to develop a predictive model for potential under-reporting. The multivariable adjusted results confirm historical prevalence of smoking as the most prominent risk factor for current lung cancer incidence. Previous literature suggests that smoking is linked to 85 per cent of lung cancer incidence (Nishida and Kudo, 2013). We used the global prevalence of age standardised smoking in 1980 (estimated to be 25.9%) as our primary predictor of incidence (Ng *et al*., 2014). This lagged prevalence was based on the assumption of an approximate 30-year latency period between exposure and incidence lung cancer rates (Gandini *et al*., 2008; Weiss, 1997).

The multivariable (adjusted) results also indicated a higher risk of lung cancer that was related to indoor pollution which had a positive association with the two variables that has been found in a majority of studies related to use of coal stoves (Lissowska *et al*., 2005; Mumford *et al*., 1987) and passive smoking (Hirayama, 2000). The association, however, excluded the high prevalence of indoor pollution in Africa due to wood and coal stoves because of the presumed high level of under-reporting of lung cancer incidence in this region (Ferlay *et al*., 2010; Lam *et al*., 2004).

The bivariate (unadjusted) results confirmed a highly significant association between outdoor pollution (PM_{10}) that has also been suggested as a potential trigger for lung cancer (Lewtas, 2007; Raaschou-Nielsen *et al*., 2013; Youlden *et al*., 2008). This particular kind of air pollution could be influenced by a large range of pollutants that includes radon, asbestos, chromium, cadmium and arsenic, organic chemicals, radiation and coal smoke. A recent study has also demonstrated the potential of remote sensed and meteorological data products as being reliable representation of global observations of $PM_{2.5}$ for epidemiological studies (Lary *et al*., 2014). Our bivariate results also indicated a positive association between GDP and lung cancer incidence illustrating the fact that rising affluence in developing countries may possibly be associated with an increase in smoking (Ezzati *et al*., 2005). Developing countries, moreover, now account for 80% of all smokers and lung cancer is projected to increase markedly in these countries (Mathers and Loncar, 2011; Murray, 1997; Tesfaye *et al*., 2007).

Predictions from our model suggest that the true global incidence of lung cancer in 2012 may have been at least 23.6 per 100,000 compared to reported estimates of 23.1 per 100,000 resulting in potentially ~38,101 missed cases of lung cancer. The high levels of under-estimation reflect the need to improve the quality of cancer incidence data noted in the data and methods section of the Globocan 2012 report. In the recent Globocan 2012 report, 62 out of 184 countries had no cancer incidence data available and only 66 reported the availability of high-quality data (WHO-IARC, 2014). The estimated incidence rate of lung cancer in many countries with poor or absence of data relies on a wide range of different algorithms (Curado *et al*., 2007; Forman *et al*., 2013). The countries lacking data were distributed among the WHO regions as follows: Africa (19), the Americas (17), Europe (13), South East Asia (5), Eastern Mediterranean (4) and Western Pacific (4). A majority of the 62 countries that had no data were also developing countries whose resources for cancer management programmes are likely to be limited. Our results reflect low-low levels of incidence clustering in Africa, which may explain the under-estimation of incidence in countries in this region if neighbouring country data were used to extrapolate incidence.

Further analysis of the 10 countries with the highest numbers of absolute missed lung cancer cases suggested that many them were African or Central-South American countries, where high levels of poor or lacking cancer data were to be found (Curado *et al*., 2007; Forman *et al*., 2013). Interestingly, separate projections of lung cancer incidence in Africa that indicate lung cancer will be a leading cause of mortality by 2020 may contradict the low Globocan 2012 incidence levels in many African countries (Mathers and Loncar, 2011; Murray, 1997). Another concern is that the suggested high levels of under-reporting of lung cancer incidence in developing countries like Lesotho, Sierra Leone, Namibia, Madagascar, Rwanda and Burundi is projected against a backdrop of limited healthcare resources (Narasimhan *et al*., 2004). A recent study in 27 EU countries suggests that the costs of lung cancer for 2008 was EUR 18.8 bn and that a considerable portion of this expense had to be covered by friends and relatives (Edge *et al*., 2010). The full cost of lung cancer in the EU is not only under-estimated, but also currently requires substantial inputs by friends and relatives (Jemal *et al*., 2011). This has implications for developing regions like Africa where a rising incidence of the disease is projected in coming years (Youlden *et al*., 2008).

Based on a recent study from Northern Africa (Tachfouti *et al*., 2012) which attempted to cost the treatment of early and advanced stage lung cancer for the first year after diagnosis, the above mentioned missed cases at the global level could have amounted to a ranged estimated cost of ~US$ 175,264,600 (95% CI US$: 131,049,400-219,479,800) provided an assumed average early stage cost of US$ 4600 per patient (worst case scenario) and US$ 130,305,420 (95% CI US$: 97,432,380-163,178,460) if all assumed average late stage cost of US$ 3420 per patient (best case scenario from cost perspective). These estimates would be far greater if first world treatment costs were factored in as well as indirect costs (*e.g.* social and lost productivity).

Limitations of the study

Given the ecological study design, the potential for ecological fallacy cannot be discounted. As this is a secondary data analysis, missing or incorrect incidence and lagged determinant data could potentially impact the model and our projections. The potential limitations of the Globocan data are discussed in more detail (WHO-IARC, 2014). National cancer reporting systems vary by country, and data quality may thus have varied between regions. Multivariable models allow taking into account confounding factors (Benichou, 2001) and the model needs to be as complete as possible. We cannot therefore discount that potentially important missing determinants of lung cancer may have confounded our findings. Poor quality of the included determinants could also have impacted the estimated number of missed cases in particular countries, leading to over or under-estimation of lung cancer incidence.

Conclusions

The WHO Framework Convention on Tobacco Control should continue its work and expand the comprehensive implementation of the framework to all 168 countries that ratified this policy in March 2010 (Tesfaye *et al*., 2007). In particular, the results suggest significant clusters of high-high male and female incidence across neighbouring countries that could provide synergies for WHO regional policy development. Nevertheless, changing inter-country and gender patterns will pose differential levels of future risk for lung cancer requiring additional resource requirements.

Global environmental policies to reduce emissions continue to be disabled by a lack of participation from the world's leading polluters including North America and China. The signatory parties of the Kyoto Protocol are increasingly being questioned because of the absence of binding resolutions to reduce emissions that are fair to all. International and national health agencies should be urged to geographically target countries that reflect under-reported incidence and develop and/or strengthen programmes, as well as establish cancer registries to prioritise interventions given a limitation of resources in many developing countries (Edge *et al*., 2010; Tesfaye *et al*., 2007). The high-high clusters of countries showing an elevated level of incidence in our results suggest regional efficiencies could be obtained by international health bodies if they coordinate programmes for these clusters. Blocks of countries in Central-South America, Asia and Africa, for instance, could be targeted in order to develop synergistic interventions to both manage the disease, as well as better implement programmes like the WHO Framework Convention on Tobacco Control.

References

Allison PD, 1978. Measures of inequality. Am Sociol Rev 43:865-80.

Anselin L, 1996. Local Indicators of Spatial Association - LISA. Geogr Anal 27:93-115.

Anselin L, Ibnu S, Youngihn K, 2006. GeoDa: an introduction to spatial data analysis. Geogr Anal 38:5-22.

Assuncao RM, Reis EA, 1999. A new proposal to adjust Moran's I for population density. Statistics in Medicine 18:2147-62.

Benichou J, 2001. A review of adjusted estimators of attributable risk. Stat Methods Med Res 10:195-216.

Bonjour S, Adair-Rohani H, Wolf J, Bruce N, Mehta S, Prüss-Ustün A, Lahiff M, Rehfuess E, Mishra V, Smith K, 2013. Solid fuel use for household cooking: country and regional estimates for 1980-2010. Environ Health Persp 121:784-90.

Bruce N, Perez-Padilla R, Albalak R, 2000. Indoor air pollution in developing countries: a major environmental and public health challenge. B World Health Organ 78:1078-92.

Burbank F, 1972. US lung cancer death rates begin to rise proportionately more rapidly for females than for males: a dose-response effect? J Chronic Dis 25:473-9.

Cohen AJ, Ross Anderson H, Ostro B, Pandey KD, Krzyzanowski M, Künzli N, Gutschmidt K, Pope A, Romieu I, Samet JM, 2005. The global burden of disease due to outdoor air pollution. J Toxicol Env Heal A 68:1301-7.

Curado M, Edwards B, Shin H, Storm H, Ferlay J, Heanue M, Boyle P, 2007. Cancer incidence in five continents. Vol. IX. IARC, Lyon, France.

Deaton A, 2013. What's wrong with inequality? Lancet 381:363.

Devesa SS, Silverman DT, Young JL, Pollack ES, Brown CC, Horm JW, Percy CL, Myers MH, McKay FW, Fraumeni JF, 1987. Cancer incidence and mortality trends among whites in the United States, 1947-84. J Natl Cancer I 79:701-70.

Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, 2010. AJCC cancer staging manual. Springer, New York, NY, USA.

Ezzati M, Vander Hoorn S, Lawes CM, Leach R, James WPT, Lopez AD, Rodgers A, Murray CJ, 2005. Rethinking the "diseases of affluence" paradigm: global patterns of nutritional risks in relation to economic development. PLoS Med 2:e133.

Farnsworth N, Fagan SP, Berger DH, Awad SS, 2004. Child-Turcotte-Pugh *vs* MELD score as a predictor of outcome after elective and emergent surgery in cirrhotic patients. Am J Surg 188:580-3.

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM, 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 127:2893-917.

Finkelstein MM, 1991. Use of "time windows" to investigate lung cancer latency intervals at an Ontario steel plant. Am J Ind Med 19:229-35.

Forman D, Bray F, Brewster D, Gombe Mbalawa C, Kohler B, Piñeros M, Steliarova-Foucher E, Swaminathan R, Ferlay J, 2013. Cancer incidence in five continents. Vol. X. IARC, Lyon, France.

Gandini S, Botteri E, Iodice S, Boniol M, Lowenfels AB, Maisonneuve P, Boyle P, 2008. Tobacco smoking and cancer: a meta analysis. Int J Cancer 122:155-64.

Hirayama T, 2000. Non-smoking wives of heavy smokes have a higher risk of lung cancer: a study from Japan. B World Health Organ 78:940-2.

Islami F, Torre LA, Jemal A, 2015. Global trends of lung cancer mortality and smoking prevalence. Transl Lung Cancer Res 4:327.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D, 2011. Global cancer statistics. CA-Cancer J Clin 61:69-90.

Lam W, White N, Chan-Yeung M, 2004. Lung cancer epidemiology and risk factors in Asia and Africa state of the art. Int J Tuberc Lung D 8:1045-57.

Lary DJ, Faruque F, Malakar N, Moore A, Roscoe B, Adams Z, Eggelston Y, 2014. Estimating the global abundance of ground level presence of particulate matter (PM2.5). Geospat Health 8:611-30.

Lewtas J, 2007. Air pollution combustion emissions: characterization of causative agents and mechanisms associated with cancer, reproductive, and cardiovascular effects. Mutat Res-Rev Mutat 636:95-133.

Lissowska J, Bardin-Mikolajczak A, Fletcher T, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, Fabianova E, Cassidy A, Mates D, Holcatova I, 2005. Lung cancer and indoor pollution from heating and cooking with solid fuels. The IARC international multicentre case-control study in Eastern/Central Europe and the United Kingdom. Am J Epidemiol 162:326-33.

Lortet-Tieulent J, Soerjomataram I, Ferlay J, Rutherford M, Weiderpass E, Bray F, 2014. International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women. Lung Cancer 84:13-22.

Mathers C, Loncar D, 2011. Updated projections of global mortality and burden of disease, 2002-2030: data sources, methods and results. World Health Organization, Geneva, Switzerland.

Minagawa M, Ikai I, Matsuyama Y, Yamaoka Y, Makuuchi M, 2007. Staging of hepatocellular carcinoma: assessment of the Japanese TNM and AJCC/UICC TNM systems in a cohort of 13,772 patients in Japan. Ann Surg 245:909.

Mumford J, He X, Chapman R, Harris D, Li X, Xian Y, Jiang W, Xu C, Chuang J, 1987. Lung cancer and indoor air pollution in Xuan Wei, China. Science 235:217-20.

Murray CJLLAD, 1997. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. Lancet 349:1498.

Naghavi M, Wang H, Lozano R, Davis A, Liang X, Zhou M, Vollset SEV, Abbasoglu Ozgoren A, Norman RE, Vos T, 2015. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 385:117-71.

Narasimhan V, Brown H, Pablos-Mendez A, Adams O, Dussault G, Elzinga G, Nordstrom A, Habte D, Jacobs M, Solimano G, 2004. Responding to the global human resources crisis. Lancet 363:1469-72.

Ng M, Freeman MK, Fleming TD, Robinson M, Dwyer-Lindgren L, Thomson B, Wollum A, Sanman E, Wulf S, Lopez AD, 2014. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. J Am Med Assoc 311:183-92.

Nishida N, Kudo M, 2013. Recent advancements in comprehensive genetic analyses for human hepatocellular carcinoma. Oncology 84(Suppl.1):93-97.

Pickle LW, Mason T, Howard N, Hoover R, Fraumeni J, 1987. Atlas of US cancer mortality among whites: 1950-1980. US Department of Health and Human Services, Public Health Service, National Institutes of Health, Washington, DC, USA.

Polednak AP, 1974. Latency periods in neoplastic diseases. Am J Epidemiol 100:354-56.

Raaschou-Nielsen O, Andersen ZJ, Beelen R, Samoli E, Stafoggia M, Weinmayr G, Hoffmann B, Fischer P, Nieuwenhuijsen MJ, Brunekreef B, Xun WW, Katsouyanni K, Dimakopoulou K, Sommar J, Forsberg B, Modig L, Oudin A, Oftedal B, Schwarze PE, Nafstad P, De Faire U, Pedersen NL, Ostenson CG, Fratiglioni L, Penell J, Korek M, Pershagen G, Eriksen KT, Sørensen M, Tjønneland A, Ellermann T, Eeftens M, Peeters PH, Meliefste K, Wang M, Bueno-de-Mesquita B, Key TJ, de Hoogh K, Concin H, Nagel G, Vilier A, Grioni S, Krogh V, Tsai MY, Ricceri F, Sacerdote C, Galassi C, Migliore E, Ranzi A, Cesaroni G, Badaloni C, Forastiere F, Tamayo I, Amiano P, Dorronsoro M, Trichopoulou A, Bamia C, Vineis P, Hoek G, 2013. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). Lancet Oncol 14:813-22.

Shorrocks AF, 2013. Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. J Econ Inequal 11:99-126.

StataCorp, 2011. Stata statistical software: release 12. StataCorp LP, College Station, TX, USA.

Tachfouti N, Belkacemi Y, Raherison C, Bekkali R, Benider A, Nejjari C, 2012. First data on direct costs of lung cancer management in Morocco. Asian Pac J Cancer Prev 13:1547-51.

Tesfaye F, Nawi N, Van Minh H, Byass P, Berhane Y, Bonita R, Wall S, 2007. Association between body mass index and blood pressure across three populations in Africa and Asia. J Hum Hypertens 21:28-37.

Weiss W, 1997. Cigarette smoking and lung cancer trends. A light at the end of the tunnel? Chest 111:1414-6.

WHO, 2009. Global health observatory data repository. World Health Organization, Geneva, Switzerland. Available from: http://apps.who.int/gho/data/node.main

WHO, 2012. WHO household energy database. World Health Organization, Geneva, Switzerland. Available from: http://www.who.int/indoorair/health_impacts/he_database/en/index.html

WHO-IARC, 2014. GLOBOCAN: estimated cancer incidence, mortality, and prevalence worldwide in 2012. Available from: http://globocan.iarc.fr/Default.aspx

World Bank, 2013. World development indicators (WDI). Available from: http://data.worldbank.org/indicator

Youlden DR, Cramb SM, Baade PD, 2008. The international epidemiology of lung cancer: geographical distribution and secular trends. J Thorac Oncol 3:819-31.