Economic growth and cancer incidence

ABSTRACT

Why do we observe increasing rates of new cancer cases? Is the increasing burden of cancer mainly the outcome of higher life expectancy and better life conditions brought about by economic development? To what extent do environmental degradation and changes in life-styles play a relevant role? To answer these questions, we empirically assessed the relationship between per capita income and new cancer cases (incidence) by using cross-sectional data from 122 countries

We found that the incidence rate of all-sites cancer increases linearly with per capita income, even after controlling for population ageing, improvement in cancer detection, and omitted spatially correlated variables. If higher incidence rates in developed countries were merely due to those factors, and not also to life-styles and environmental degradation, we would have found a flat or even an inverted-U pattern between per capita income and cancer incidence.

The regression analysis was applied also to the eight most common site-specific cancers. This confirmed the existing evidence on the different patterns in rich and poor countries, explained the pattern of the estimated relationship for aggregate cancers, and gave some other interesting insights.

KEYWORDS: Economic development; Cancer; Environmental Kuznets Curve; Environmental degradation; Spatial error models.

Highlights

- New cancer cases increase with p.c. income in a cross-section of 122 countries.
- Improved detection potential and a longer life alone cannot explain this evidence.
- Bad life-styles and environmental degradation play a relevant role.

1. Introduction

Cancer incidence (yearly new cases of cancer) is increasing and predicted to grow fast. The term 'Cancer epidemic' has become frequently used, not only by the media (e.g. Servan-Schreiber, 2008), but also by academic journals and by the World Health Organization[1]. The

[1] In April 2015, the Lancet Oncology and The Lancet launched a joint campaign against cancer "to inform strategies to control the global cancer epidemic" (see http://www.thelancet.com/campaigns/cancer). In 2005 the term 'epidemic' was used in the 58th resolution of the WH assembly, see http://www.who.int/mediacentre/news/releases/2005/pr_wha05/en/

problem is particularly alarming in lower- and middle-income countries (see, e.g., Boyle and Levin 2008; GLOBOCAN 2012; Stewart and Wild 2014; Vineis and Wild 2014; Ferlay et al. 2015; Torre et al. 2015). For some rich countries, incidence rates are stabilizing (or slightly decreasing), however at very high rates. In the USA, this has been the case since the mid 1990s (Siegel et al. 2016).

Although data availability on cancer has increased significantly in the last years[2], the relationship between cancer incidence and economic development remains largely unexplored, with just a few exceptions, namely: Beaulieu et al. 2009, Bray et al. 2012, Fidler et al. 2016.[3] The first is a report by "The Economist" Intelligence Unit on the health and economic burden of cancer. As a supplementary result, in one of its appendices, the report shows the outcome of a multiple regression analysis aimed at understanding cross-country variations in both estimated cancer incidence rates for 2009, and in fatality rates for 2002. Regressors included p.c. income, per cent of population aged 65+, and regional dummies. The authors found a positive association of higher cancer incidence rates with both age and higher per capita income countries, which they attributed to the belief of "underreporting of cancer cases in developing countries" (Beaulieu et al., 2009, 62).

Bray et al. (2012) and Fidler et al. (2016) grouped countries according to the four levels (low, medium, high, and very high) of the Human Development Index (HDI) and compared incidence and mortality rates across groups. Both articles brought support in favour of the so-called "cancer-transition", according to which the demographic transition and economic development are changing the composition of the different types of cancers, with a shift from cancers linked to infections to those associated with non-infectious risk factors and possibly associated with the "western" lifestyle.

The above-mentioned papers are in line with the health literature, briefly summarised in the next section. The general idea is that increasing cancer incidence rates might be the outcome of economic development, which delivered not only higher life expectancy and improved cancer detection and statistical reporting, but also environmental degradation and "bad" life-styles.

The aim of our research was to empirically investigate the macro level relationship between cancer incidence rates and per capita income. For this purpose, we tested some reduced models that looked only at the ends of the complicated causal chains. Such an approach has been followed by the so-called Environmental Kuznets Curve (EKC) literature

52 For an assessment of the status of population-based cancer registries worldwide see (Bray et al., 2015).
63 The differences between the present research and the previously mentioned studies will be discussed in section 5.

that has been investigating the relationship between economic growth and the environment for more than 25 years (e.g., Stern 2004, Dinda 2004, Luzzati 2015). While the EKC literature focused on anthropic pressures, e.g. emissions, here we focused on one possible outcome of pressures, that is, cancer occurrence.

The paper is structured as follows. The second section outlines the links between cancer and economic development, from which we derived the conceptual model for our empirical analysis (Figure 1). The third section describes data and methods. In the fourth section results are presented and discussed. The last section gives our conclusions.

2. Cancer and its possible links with economic development

This section firstly summarises what we know about cancer genesis, and then why economic development can play a major role in cancer occurrence. The dominant theory explaining cancer is the so-called Somatic Mutation Theory (SMT) (Nowell 1976; Hanahan and Weinberg 2000 and 2011) according to which "random mutations in the genes which control proliferation or apoptosis are responsible for cancer" (Bertram 2001, p. 170). Hence, cancer is due to stochastic (relevant) mutations that occur in oncogenes and tumour suppressor genes (Lodish et al. 2000). The older a person, the higher is the number of accumulated stochastic mutations, which ultimately leads to higher probability of cancer occurrence.

Recently, SMT has been criticised on the basis of theoretical reasons and experimental and epidemiological evidence. Hence, other theories of carcinogenesis have begun to gain ground. They shift the focus from single cells to the entire tissue and attribute a prominent role to altered environments (epigenetic signals) for regulating gene expression, rather than to stochastic mutations of DNA (see e.g. Burgio and Migliore, 2015). For instance, Tissue Organisation Field Theory (TOFT) (see e.g. Baker 2015), which is better seen as integrative rather than alternative to SMT (Bedessem and Ruphy, 2015), looks promising for understanding the role of low-dose foetal exposure to ubiquitous and long lived chemical pollutants, namely the endocrine-disrupting chemicals (EDCs)[1]. These chemicals, by mimicking physiologic hormone signalling molecules, perturbate tightly regulated intercellular signalling pathways. This leads to subtle architectural changes in tissue organization that increase the risk of cancer development. (Howard and Stats 2013).

Overall, cancer is increasingly seen as the disruption of a complex equilibrium, that is, the outcome of an evolutionary process in which random genetic mutations have to face the selection of environmental pressure; moreover, intrinsic epigenetic plasticity, clonal evolution

81 A useful introduction to EDCs is Gore et al. (2014).

and high cellular adaptability are also crucial (Greaves, 2014). Hence, cancer is acknowledged as stemming from many interacting factors, that is, from mutations in oncogenes and tumour suppressor genes, from genetic inheritance[2], work and living environment, and lifestyles (see e.g. Belpomme et al. 2007a, Belpomme et al 2007b, Stewart and Wild 2014).

Many studies have investigated the differential contribution to cancer incidence of non-genetic risk factors (e.g. Danaei et al. 2005) and of environmental factors (e.g., Alavanja 2003, Boffetta 2006, Mannucci et al. 2015, Stare and Jozefowicz 2008). The confluence of diverse types of evidence increasingly indicates the relevance of involuntary exposure to environmental contaminants, which affect particularly the "developing foetus, the developing child and adolescent" (Newby and Howard, 2005, 57). For instance, there is evidence of decrease in the average age of cancer onset (e.g. Newby et at. 2007) and increase in childhood cancers (e.g. Steliarova-Foucher et al. 2004), which are also attributed to environmental factors (Stewart and Wild, 2014; Norman et al., 2014). Historical evidence supports the idea that cancer is a disease of industrialization/wealth since "in preindustrial societies, the death rate in infancy was high, but if adolescence was reached then [...] the chances of living a reasonable life span in good health were high and unlikely to end in the development of cancer" (Howard and Statts, 2013). It is not under dispute that economic and technological progress led to the introduction of a complex mixture of persistent xeno-chemicals and other pollutants that have been recognised as carcinogenic.

Aggregate quantifications of the environmental risk factors have been proposed in a wide–ranging report by the World Health Organization that surveys the findings on the environmental risk factors (Prüss-Üstün, 2016). According to this report, household and ambient air pollution, passive tobacco smoking, radiation, chemicals, and occupational risks are responsible for at least 20% of cancer cases (in terms of disability-adjusted life years) (Prüss-Üstün, 2016, XVI, 50, and 86).

Of course, any estimate is highly uncertain (and incomplete) because of the complexity of the cancer-environment relationship in which polluting agents are often time persistent and pervasive, bio-accumulate and are bio-magnificated along the food chain, performing multiple biological actions as well as acting in synergy with other substances. Because of this complexity, we believe that it is also useful to tackle the issue with a very coarse grained perspective, empirically investigating to what extent economic development as a whole plays a role in cancer incidence. To this end, we performed a regression analysis in which income per

[2] The heritable factors have an important, but not exclusive, role. For instance, using data from Swedish, Danish and Finnish twin registries, it has been reported (Lichtenstein et al 2000) that genetic influence on the incidence of cancer explains no more that 42% of the variance in incidence rate, depending on the cancer site.

capita proxies the joint effects of environmental factors and life-styles on cancer incidence. The reasons why income per capita is expected to be a significant regressor of cancer incidence come from analysing the major links of the causal chain that goes from income to cancer, which are described in what follows and summarised by Figure 1.

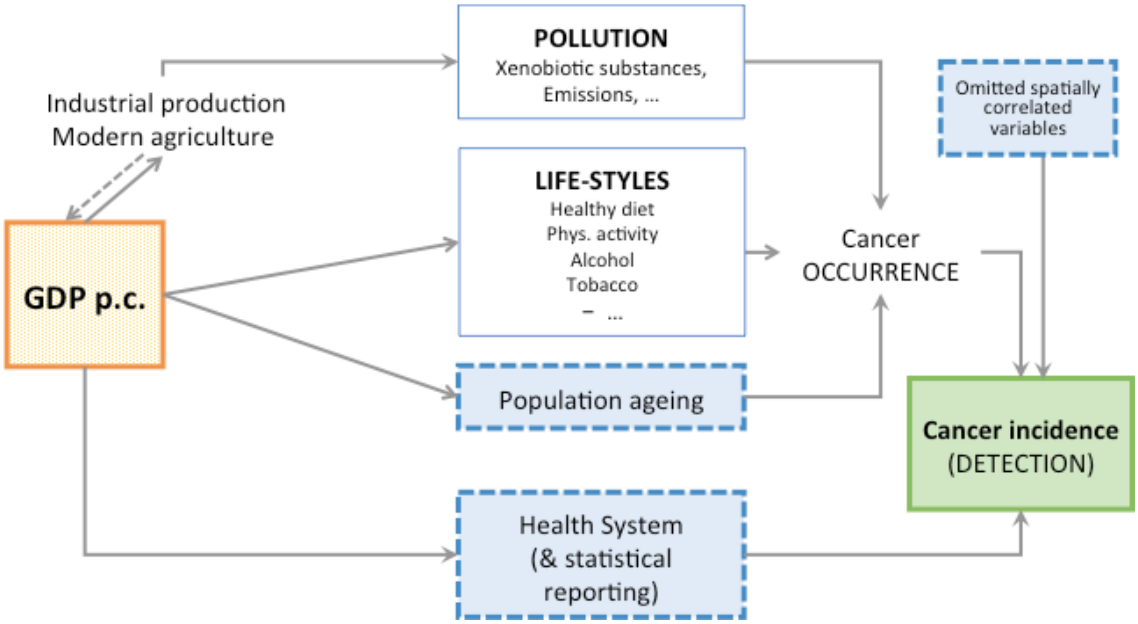


Figure 1: From income to cancer incidence: major links

Economic development started with the industrial revolution and was literally fuelled by fossil fuels (e.g., Smil 2000). The availability of an unprecedented quantity of energy radically transformed our economy and the relationship between humans and nature, to the point that many scholars believe that we entered a new epoch, the Anthropocene (Crutzen 2002, Steffen et al 2011). A large amount (and number) of pollutants have populated the places where we live, resulting in prolonged and pervasive biochemical stresses that have been found to be risk factors for several diseases, including cancer. Furthermore, other new risk factors (e.g. excessive-weight and obesity) emerged as an outcome of life-style changes that have accompanied the economic development process.

At the same time, material living conditions have generally improved, thereby on the one hand bringing about an increase in cancer due to higher life expectancy and, on the other, leading to a reduction in cancers related to some infectious diseases. In other words, income growth has allowed an epidemiological transition[1], that is, a shift "from a predominance of

[1] According to the theory of epidemiological transitions (Omran 2005, pp. 737-738), three ages of mortality patterns in history are observed, namely the age of "pestilence and famine", of "receding pandemics", and of "degenerative and man-made diseases". In the first "age" life expectancy at birth is very low, but epidemic peaks then become less frequent or disappear,

cancers linked to infections to cancers associated with risk factors that are mainly non-infectious and possibly related to the so-called western lifestyle" (Maule and Merletti 2012, p. 745). The identification of this "new epidemiological age" is not only a theoretical construct, but also a relevant empirical fact. According to the World Health Organization (WHO, 2014) about 52% of worldwide deaths in 2012 were due to Non-Communicable Diseases (NCDs) and, of these, about 27% were associated with Malignant Neoplasm.

As a concluding remark, it has to be considered that part of the observed increase in cancer incidence along the economic development is the result of improved diagnostic scrutiny and statistical reporting (e.g. Li et al. 2013, Moynihan et al. 2012). In countries where health systems are not well developed, cancer statistics collection is poorly organized and the causes of death often remain undiagnosed resulting in under reporting of cancer deaths in less developed countries (e.g. Fallah and Kharazmi, 2008).

3. Material and methods

3.1 The empirical model

The regression model used in this paper is visualised in Figure 1. The items in the dashed contoured boxes have been controlled for in the regressions, so that the variability of incidence rates explained by income can be seen as coming from the joint effect of lifestyles and pollution. The design of the present analysis does not allow a distinction between lifestyles and environmental risks, the importance of which, however, can be drawn from the health literature that was briefly summarised above.

Building on the arguments put forward in the previous section, we regressed cancer incidence rates on the 20-year lagged p.c. income while controlling for (1) population ageing, (2) potential for detection and statistical reporting, and (3) omitted factors that might be related to the country's geographic location.

Lags in income were used to take into account the long genesis of cancer and its possible epigenetic nature (see, e.g., Burgio and Migliore, 2015). There are no strong theoretical reasons for taking a particular time lag. For instance, there seems to be a 30-35 years lag between the peak in tobacco consumption and the peak in the fatality cases of lung cancer attributable to tobacco smoking (Stewart and Wild 2014, p. 82 ff.; Bilano et al. 2015). This lag is consistent with a lag of 20 years, or more, when considering cancer occurrence. We chose a 20 year lag since it was the longest available time period, due to some lack of older

after which we eventually enter a phase in which mortality tends to approach stability at relatively low levels and non-communicable diseases, including malignant neoplasms, prevail.

187data for income. In any case, we also checked for different time lags (none, 5, 10, and 15
188years), finding that results do not change qualitatively. This should not be surprising since
189income is highly autocorrelated. Hence, from an empirical point of view, the choice of the time
190lag has low relevance. In any case, using lagged values for income is important from a
191statistical point of view since it avoids potential endogeneity issues.

192 To control for population ageing, we used average standardised rates since they take
193the different age profiles of countries into account (see below). Furthermore, given that the
194small size of older age classes in poor countries could cause incidence rates to lose statistical
195significance, we did a further check by analysing the age class 40-60 separately.

196 Improvements in cancer detection and statistical reporting along the process of
197economic development were proxied by physician density (physicians per 1000 inhabitants).
198The reasons for choosing this variable are discussed in detail in the next section.

199 Many other potential factors (such as genetic risk or diet and habits) can be considered
200as strongly related to the geographic location of the country. Those factors have been omitted
201since they are either unobservable or lacking in reliable data. A spatial error model was used
202to take into account these omitted spatially correlated covariates.

203<u>3.2 Estimation methods and techniques</u>

204 In order to choose the model that best fits the data, papers within the EKC literature
205often compare parametric estimates with different specifications of the p.c. income term, i.e.,
206linear, quadratic or cubic (see Van Alstine and Neumayer, 2010). We followed a different
207approach. As in Luzzati and Orsini (2009) we preliminarily used semi-parametric methods to
208assess whether a linear or non-linear specification better fits the data. In the case of evidence
209of a linear fit, we used it in the parametric estimates. If semiparametric fits suggested non-
210linear patterns, we chose between the quadratic and cubic specification by minimising the
211corrected Akaike Information Criterion (AICc) (which also involved maximisation of the
212adjusted R-squared).

213 For the semiparametric estimates, we used the generalized additive model (GAM), in
214which each variable enters nonlinearly and separately. We followed the approach proposed in
215Wood (2006), which is based on penalized regression splines, and used the “mgcv package” in
216R Development Core Team (2012), with the restricted maximum likelihood (REML) option
217(see Wood 2011).

218 For the parametric estimates, we followed the spatial econometric methodology
219developed by Anselin (1988). The reason behind this choice is that, differently from the

220inclusion of regional dummy variables (as e.g. in Beaulieau et al. 2009), this methodology
221explicitly accounts for the effects of spatial correlation due to imperfections in model and
222measurements that exhibit a spatial structure, thereby increasing the efficiency of the
223parameter estimates.

224 The spatial correlation can be incorporated in a regression model in different ways. In
225the current paper, we used the spatial error model (SEM) in which the spatial correlation is
226modelled in the error term. This is based on the *a priori* grounds that cultural and genetic
227factors vary across space and are assumed to drive cancer incidence variability. However, due
228to lack and/or unreliability of data, such factors are unobserved and/or unmeasured and,
229therefore, omitted in the regression. If they are influential, then their impact on the
230explanation of cancer incidence is subsumed in the error term that shows a spatial pattern.

231 To check if this is the case, a spatial specification search was carried out using OLS
232estimation and applying a Lagrange Multiplier test. In the presence of evidence of spatial
233correlation, the spatial model was estimated by means of maximum likelihood.

234 The SEM was implemented by specifying a spatial stochastic process for the error term,
235which in turn yields the nonzero correlation for the units that are considered as neighbours.
236Consequently, the spatial error model requires the definition of a spatial matrix, which reflects
237the potential interactions between neighbouring units (countries in our case). Here, two
238different countries are considered as interacting with each other if and only if they belong to
239the same region. The region taxonomy was taken from the International Agency for Research
240on Cancer, the specialized cancer agency of the World Health Organization (WHO), and is
241listed in the appendix. The spatial regressions are estimated using the "spdep package" in R
242Development Core Team (2012).

243<u>3.3. Variables</u>

244 Data on cancer incidence are becoming increasingly reliable due to the diffusion of
245national cancer registries (see, e.g., Parkin and Donald 2009). However, cross-national
246differences in coverage and quality of the data collected are quite pronounced, resulting in
247high variability of both coverage and reliability: thus the quality is often associated with the
248level of economic development. For a worldwide comparison, the most relevant project is
249GLOBOCAN, which is today incorporated in "Cancer today"[1]. GLOBOCAN is a project of the
250International Agency for Research on Cancer of the World Health Organization. Its database
251contains data for 26 site-specific cancers and for all sites cancer (excluding non-melanoma

221 See http://globocan.iarc.fr/Default.aspx

skin cancer). This project produced the most recent (2012) estimates of incidence, mortality and prevalence.

In order to control for differences arising merely from differences in the age profiles of each population, the average standardized rates (weighted) - ASR(W) - have to be used. The standardization procedure (for details see, e.g., Boyle and Parkin 1991) adjusts observed age-specific rates to a reference population, commonly referred to as the Standard Population, usually the world population. The term 'weighted' refers to standard weights taken from the population adopted as a standard. We calculated[2] ASR(W) using the population weight of the World Standard Population[3] and the population data of the United Nations database.

The p.c. income variable was the p.c. GDP, expressed in thousands of US$ PPP2011[4], taken from the World Bank online database. Income was averaged over three years to mitigate the effect of the business cycle. As stated previously, we used a 20 year lag to consider the long genesis of cancer and tested shorter income time lags, which however left results qualitatively unchanged, as one would expect from the strong autocorrelation of p.c. income (see the appendix, Table A2).

As regards the variable to proxy the diagnostic potential of a country, it has to be emphasised that detection and statistical reporting are very different from early detection. While the first two affect the quality of the incidence rates data, the latter is relevant for treating cancer, and hence for mortality rates. Early detection is strongly associated with the presence of screening programmes and diagnostic facilities, which are in turn associated with high levels of per capita health expenditure (and income). For mere detection and statistical reporting, however, an easy access to a doctor is a crucial variable, more important than the availability of advanced technical tools. So far, physician density has proved to be very important in cancer detection (e.g., Ananthakrishnan et al. 2010, Fleisher et al. 2008, Li et al 2013, Sundmacher and Busse 2011) and for many other care issues, like infant mortality (e.g., Farahani et al. 2009), and generally for health outcomes (e.g. Friedberg et al. 2010, Macinko et al. 2007, Mondal and Shitan 2014, Shi 2012). At the same time, physician density can be reasonably thought as having "diminishing returns" in cancer incidence reporting, that is, after

2 The database provided by GLOBOCAN already provides ASW(R) rates. Using the data available online and implementing the procedure described by the Glossary section of GLOBOCAN 2012 (http://globocan.iarc.fr/Pages/glossary.aspx) we obtained slightly different figures.

3 http://seer.cancer.gov/stdpopulations/world.who.html World Standard Population is used also in GLOBOCAN 2012.

4 GDP was taken in Power Purchasing Parity (PPP2011) due to the cross-country nature of the analysis. PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as a U.S. dollar has in the United States. Data are in current international dollars based on the 2011 International Comparison Program (ICP).

some thresholds, further increases in the physician density will have increasingly smaller relevance.

For the above reasons, we took physician density as a proxy of cancer incidence reporting potential and used it in the regressions with a concave specification. This was empirically supported by the positive and decreasing marginal impact of physician density on cancer incidence in preliminary semi-parametric estimates (see, e.g., Figure A2). Data for physician density (physicians per 1000 inhabitants) were taken from the World Bank online database. But for a few exceptions, they range from 2010 to 2012. The correlation between physician density and GDP p.c. is not strong enough to prevent the use of both variables as regressors (see Table A2).

We avoided transforming the variables into logarithms since this practice, although common, has been shown to be theoretically weak (Mayumi and Giampietro, 2010). Nonetheless, we verified that using logs does not change the results qualitatively.

3.4. Countries

The GLOBOCAN 2012 dataset covers 184 Countries. We excluded those countries (33) for which data were estimated by merely imputing the data of neighbouring countries or registries in the same area. Of the 151 remaining countries, we excluded 5 that are not included in the World Bank online database from which we took per capita income[1]. We also excluded 18 countries for which 20 year lagged p.c. income or other data were not available. Our final list, presented in the appendix, included 122 countries since six other countries were considered outliers and removed.

As discussed in econometrics textbooks (e.g. Gujarati 2004, 540 ff.), including or excluding outliers is a tricky issue. An outlier differs markedly from the other observations and, hence, "provides a large residual when the chosen model is fitted to the data" (Draper and John 1981, 21). An outlier must be excluded if it is influential, that is, distorting the slope of the regression line or even forcing the researcher to change the model specification. This problem is particularly serious in semiparametric models since "GAMs can be very sensitive to the presence of a small proportion of observations that deviate from the assumed model. In other words: a few atypical observations could seriously affect the non-parametric estimates of the smooth regression function" (Azadeh and Salibian-Barrera, 2011).

A preliminary check on the dataset (see Figure A1) shows that there are some observations that are potentially influential (due to their very high income levels) and for which one can easily imagine that they will have large regression residuals for any

[1] State of Palestine, France Guadeloupe, France La Reunion, France Martinique, and France Guyana.

specification that can be conceived. One notices that the observations in question refer to two very small and atypical countries, Singapore (a city-state with a rather idiosyncratic economy) and Luxembourg (whose economy is based on financial services), and to another 5 countries whose economy is strongly based on oil exports[2]. Their special characteristics are such that other countries cannot be thought to mimic their performances. For a formal check of influential observations and outliers we followed the approach developed by Fox (2008)[3], which is based on studentized residual, hat-values and Cook's distances. Applying it to several model specifications (the linear model and those described in section 4), we found that the 5 "oil" countries and Luxembourg should be excluded, while excluding Singapore is not statistically supported.[4]

3.5 Data descriptive statistics

A preliminary overview of the data is given by Table A1, which contains the main descriptive statistics for the variables. Table A2 shows the correlation matrix for all-sites cancers, both for the entire population and for the age class 40-60, p.c. income (and its lagged values), and physician density. As expected the autocorrelation of p.c. income is remarkably high.

4. Results

We start by presenting the results for all cancers, and then we move to organ site-specific cancers. The labels of the variables are as follows. *AllC* refers to incidence rates for all cancers, otherwise the name of site-specific cancer is indicated. The suffix "_40-60" indicates that the rate refers to the population in the age class 40-60. Incidence rates are measured as yearly new cases on 100,000 inhabitants. *Y_92* is the three-year average, centred on the year 1992, of GDP p.c. (thousands of $PPP2011) and *PhysD* is the physician density in 2012 (number of physicians every 1000 inhabitants).

The semiparametric analysis for all cancers is shown in Figures 2a and 2b. Regressing the incidence rate on p.c. income gives the concave curve that is shown in Figure 2a. However, when controlling for physician density, the marginal impact of p.c. income on cancer incidence becomes linear (the straight line in Figure 2a), while the marginal impact of physician density is non-linear (Figure A2). Similar results (not shown) are obtained when analysing the 40-60 age class. Figure 2b shows the confidence bands (5%) of the regression shown in Figure 2a.

Bahrain, Brunei, Oman, Saudi Arabia, United Arab Emirates.

The approach is also described in Levshina (2015, 153-155). We used the *influencePlot* function in the R-Package 'car' (https://cran.r-project.org/web/packages/car/car.pdf).

It is worth noticing that including the 5 "oil" countries requires changing the model specification, while Luxembourg affects only the size of the estimated coefficients.

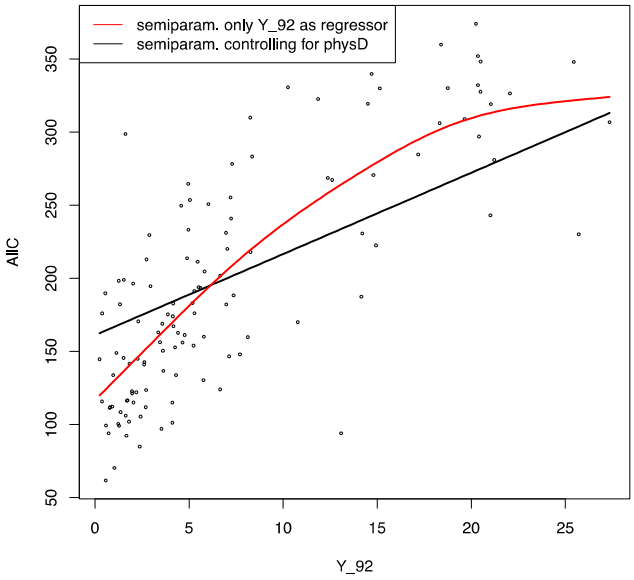


Figure 2a. Standardised cancer incidence rates in 2012 vs. p.c. income in 1992. Semiparametric fits when controlling (straight line) and not controlling (curve) for physician density. Age classes: whole population.

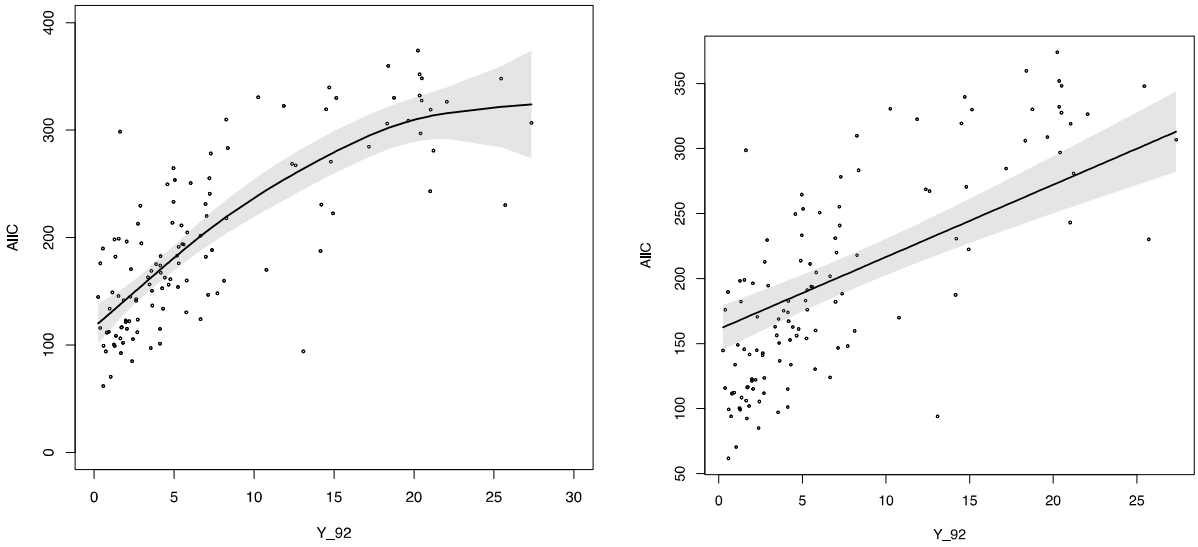


Figure 2b. Standardised cancer incidence rates in 2012 vs. p.c. income in 1992. Semiparametric fits with confidence bands (5%) when not controlling (left) and controlling (right) for physician density. Age classes: whole population.

This preliminary evidence, as discussed in detail in section 3.2, was helpful to specify the parametric estimates, which, in contrast to the semiparametric estimates, also allow the possible spatial correlation of errors to be taken into account.

The results of the SEM parametric regressions for all cancers are shown in Table 1. Let us start from the first three equations. Equations 1a and 1b refer to the whole population,

357while equation 2 concerns the age class 40-60. In equation 1a income is 20 years lagged, while 358in eq. 1b income has no lag. The two estimates are very similar. As already mentioned (see 3593.1), this is due to the high autocorrelation of income, and also holds for estimates referring to 360other time lags. All estimates show that the incidence rates for all cancers are positively 361correlated with p.c. income even when controlling for population ageing, physician density 362and omitted spatially correlated covariates. Standardised incidence rates increase by 4.66 and 363by 0.64, respectively for all ages (eq. 1a) and for the age class 40-60 (eq. 2), per increase in p.c. 364income of 1,000$ (1992, PPP2011). Figure 3 and Figure A3 show the marginal impact on 365standardised incidence rates of p.c. income, drawn respectively for the whole population (eq. 3661a) and for the 40-60 age class (eq. 2). It should be noted that p.c. income coefficients may 367partially capture differences in detection/reporting capacity that cannot be accounted for in 368terms of physician density. However, as Figure 2 suggests, there are no reasons for believing 369that including a better proxy would transform the relationship from linear to concave .

370 Table 1. Summary of the Spatial Error Model parametric estimates[1]

Dep. Var.												
AllC	=	108.8	+	4.66 *Y_92*	+	47.94	*PhysD*	-	6.09	*PhysD*^{2}		[Eq. 1a]
s.e.		9.71		0.81		8.35			1.41			
p.		<0.001		<0.001		<0.001			<0.001			
				AdjR^2=0.76; Spatial parameter=0.47, p<0.001								
AllC	=	107.06	+	2.20 ***Y_12***	+	46.06	*PhysD*	-	5. 860	*PhysD*^{2}		[Eq. 1b]
s.e.		0.60		0.34		8.06			1.36			
p.		<0.001		<0.001		<0.001			<0.001			
				AdjR^2=0.78; Spatial parameter=0.49, p<0.01								
AllC_40-60	=	40.05	+	0.64 *Y_92*	+	18.57	*PhysD*	-	2.47	*PhysD*^{2}		[Eq. 2]
s.e.		3.30		0.29		2.97			0.50			
p.		<0.001		<0.001		<0.001			<0.001			
				AdjR^2=0.58; Spatial parameter=0.41, p<0.01								
AllC	=	130.93	+	11.94 *Y_92*	-	0.24	*Y_92*^{2}					[Eq. 3]
s.e.		12.04		2.31		0.10						
p.		<0.001		<0.001		0.015						
				AdjR^2=0.72; Spatial parameter=0.58, p<0.01; calculated turning point *Y_92*=25.402								

371

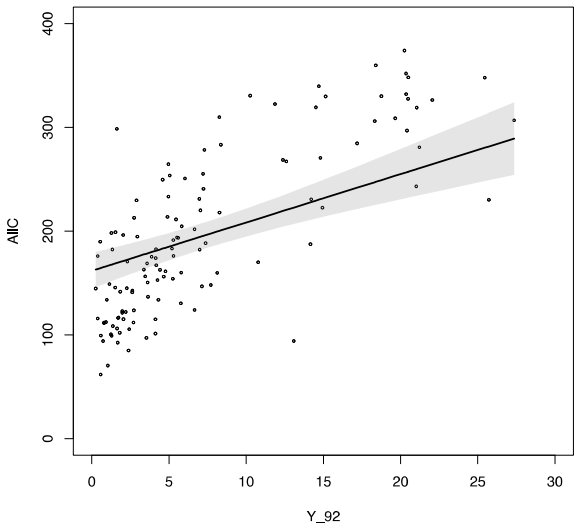
421 Results are rounded to two decimal places.

Summary statistics of the observed values:		*Min*	*Mean*	*Max*
	AllC	61.8	196.3	374.1
	AllC_40-60	18.3	64.7	115.2
	Y_92	0.251	7.266	27.352
	PhysD	0.02	1.80	6.72

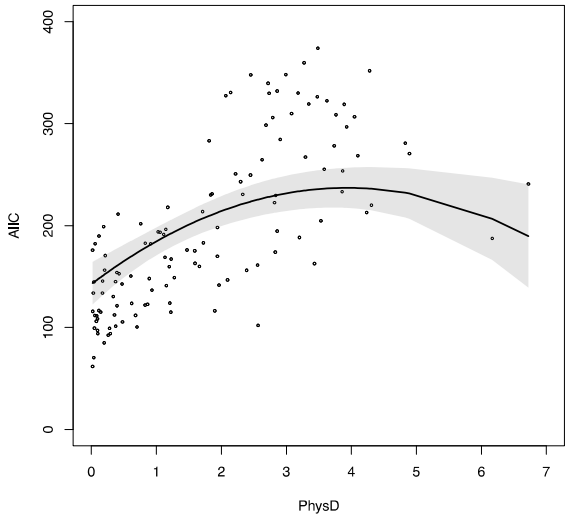
372

373 The contribution of physician density to the fitted incidence rates is measured by the 374corresponding terms of Equations 1a and 2 and can be visualised by Figure 4 and Figure A4, 375drawn respectively for the whole population and for the 40-60 age class. For both regressions, 376the impact is positive and increasing only up to roughly 3.8 physicians over 1000 inhabitants. 377Further increases in physician density beyond this value cannot be assessed since very few 378countries surpass it and the confidence bands become very wide. Similar behaviour emerged 379in all the other regressions where physician density was significant.

380



igure 3. Marginal impact (and 95% confidence band) of p.c. income.
(Parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: all population).



igure 4. Marginal impact (and 95% confidence band) of physician density.
(Parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: all population).

381

382 Equation 3 is shown only to illustrate the effect of not controlling for physician density, 383which however involves mis-specification. Given that the semiparametric fit (see Figure 2) 384was non linear, by minimizing AICc we found that in this case the best specification for 385income was quadratic. The fitted curve has an inverted-U shape. The reason is that the 386quadratic term in equation (3) partially captures that which in equation (1) is captured by 387physician density. The estimated relationship is increasing within almost the entire range of 388the observed income values since the calculated turning point (25,402$) is close to the highest

p.c. income (27,352$) (see Table 1). Not controlling for physician density yields similar consequences in all the other regressions for which the estimated relationship between income and cancer was positive.

The analysis presented above was replicated for most common site-specific cancers. Table 2 and Table A3 in the appendix, referring respectively to all age classes and to the 40-60 age group, summarise the main results. They are organised as follows. Organ site-specific cancers are ordered according to their relative frequency, which is shown in the second column. The third column indicates whether, according to the health literature (see below) the cancer organ site is typical of high-income countries (H) or low/medium income (M-L) ones. Notice that the most frequent organ-site cancers are also typical of high-income countries. The fourth column gives a concise indication of the shape of the estimated relationship while the fifth column indicates whether the estimated income coefficient is positive, negative, non-significant, or follows an EKC shape. The sixth column shows whether the estimated effect of physician density is positive or concave (+), negative (-), or non-significant (n.s). The seventh and eighth columns give the values of the income coefficients and their significance level, while the ninth shows the adjusted R-squared. Full results are in the appendix, Table A4.

The prevalence of different organ sites cancer according to the level of development is well established (e.g. Newby et al. 2005; Stewart et al 2014). Lung, breast, colorectum and prostate cancers are the most common organ sites in developed countries, associated both with lifestyles and with environmental factors (Howard and Staats, 2013). Liver, stomach, oesophagus and cervix uteri are highly correlated to chronic infection (such as hepatitis B virus, human papillomaviruses and Helicobacter pylori), which are more common in low/medium income countries.

When regressing incidence rates on GDP p.c., the sign of the coefficient should be negative for organ-sites typical of low-income countries, and positive (or EKC shaped) for cancers typical of high income countries. This is because increase in GDP in low-income countries improves overall hygiene conditions and thus reduces cancer incidence associated with chronic infection types of cancers. In contrast, in high-income countries an increase in GDP leads to an increase of environmental pollutants and xenobiotic substances and adoption of “western” type life-styles that bring an overall increase in cancer incidence (see discussion in section 2).

Our regressions confirmed the expected typicality for developed or LDC countries (see the signs of the income coefficients in Table 2). Results also highlighted that colorectum

cancer is the only one among the most frequent organ-site cancers that follows an EKC shape, although the curve turns down only at rather high levels of p.c. income (at 22,890 1992$ GDP p.c.). The external risk factors associated with this organ cancer are mainly life-style factors, namely diets that are high in red and processed meats, habitual inactivity, alcohol use, and tobacco smoking. The latter is the most important risk factor for developing lung cancer; however, several environmental and occupational exposures have also been found to be relevant, explaining globally 36% of lung cancer (Prüss-Üstün, 2016, 50). The composition of smoking with other exposures could explain the linearity of the pattern for lung cancer. The environmental risks that are positively associated with income could have offset the beneficial effects of the reduction of cigarette smoking that occurred in many high-income countries (see Stewart et al., 2014).

The colorectum and lung cases highlight the finding that differences in relative frequency between developed and less developed countries (LDCs) do not imply that environmental factors are irrelevant in the organ-site cancers typical of LDCs. For instance, ionizing radiations, exposure in the rubber industry and to asbestos are also risk factors of stomach cancer. Among the sites that we considered, the prostate, the cervix, and the liver are likewise associated with environmental factors (Prüss-Üstün, 2016, 46-51), although the first is typical of high-income countries while the latter two are most frequently found in low-income areas.

In any case, disentangling the environmental effects is difficult since, for each organ site cancer, occurrence is affected by several risk factors that differ according to income. Hence, the actual shape of relationship between income and incidence depends on the relative strength of the various contrasting effects.

Two further remarks can be made. First, physician density was found to be negatively correlated with cervix uteri cancer, which is consistent with the importance of physicians in fostering prevention, and hence reducing incidence rates of this organ site cancer. Second, the estimates for the different organ-site cancers help to understand why a positive relationship with income emerges at the overall level. The reason is that the cancers for which a negative relationship with income holds are less frequent than those for which the relationship is positive.

Table 2. Summary of the Spatial Error Model estimates for the 8 most common organ-site cancers , all age classes

1	2	3	4	5	6	7	8	9
Organ Site	*Rel. freq.*	*Typical of*	*Model*	*Role of Y_92*	*Role of PhysD*	*Y_92 coeff.*	*Y_92^2 coeff.*	*Adj RSq*

Lung	13.0%	H	Linear	+	+	0.41**		0.61
Breast	11.9%	H	Linear	+	+	1.89***		0.70
Colorectum	9.7%	H	EKC	$	+	2.25***	-0.049***	0.76
Prostate	7.9%	H	Linear	+	n.s.	2.43***		0.75
Stomach	6.8%	L/M	Linear	-	+	-0.37***		0.36
Liver	5.6%	L/M	Linear	-	n.s.	-0.36**		0.08
Cervix uteri	3.7%	L/M	Linear	-	-	-0.39^		0.48
Oesophagus	3.2%	L/M	None	none	n.s.	n.s.		0.28

455 H: High-income countries, L/M: low/medium income countries
456 Significance levels: **: p<0.05, ***: p<0.01, ^: 0.11; n.s.: non-significant
$ calculated turning point = 22,890 $ GDP 1992PPP

457

458 Finally, the estimated coefficient of the spatial dependence in the error term was
459positive and significant in all regressions, confirming the relevance of omitted geographically
460correlated factors.

461

462**5. Conclusion**

463 The evidence presented in this paper can be compared with the results of three previous
464studies, already mentioned in the introduction, the primary goal of which, however, was not to
465explore the relationship between cancer incidence and income growth. Beaulieu et al. (2009)
466used a methodology similar to ours, that is, they focused on p.c. income and performed a
467regression analysis. In contrast to us, they controlled for the effect of population ageing by
468including in the regressions the percentage of population aged 65; additionally, they used
469intercept dummies to control for geographical differences, but did not control for non-linear
470influences of income and for cancer reporting improvements. Moreover, they did not use time
471lags for income. Finally, they had to produce their own estimates of incidence rates, based on
472GLOBOCAN data for 2002, while we were able to use more reliable and recent data (2012).
473Hence, their results are not fully comparable with ours. Beaulieau and colleagues interpreted
474the positive relationship between incidence rates and income as an expected outcome of
475underreporting cancer cases in developing countries. We show that the positive relationship
476does not disappear when controlling for physician density, taken as a proxy for incidence
477reporting. On the contrary, our data and analysis suggest a higher effect of income on cancer
478incidence rates: in Beaulieau et al. (2009, 63) the income coefficient is equal to 1.457, with a
47995% confidence interval of [0.50; 2.4], while ours is equal to 4.66 (see equation 1a), with 95%
480C.I. of [3.06; 6.26]

481 Bray et al. (2012) and Fidler et al. (2016) did not use regression analysis or income as a
482key variable: rather, they compared four groups of countries pooled according to the level of
483HDI. This could be problematic because of ex-ante defined groups and because HDI also
484includes life expectancy, which should, instead, be a control variable. In any case, their design

485is too different to allow for a close comparison of the results. Nonetheless, it can be noticed 486that their papers likewise found both a positive relationship of incidence with levels of 487development, and different patterns for different cancer sites when comparing less developed 488and developed countries.

489 Overall, our results are in line with previous evidence, which is not only updated but 490also strengthened because of the use of a different methodology. Our regressions, which 491explain a substantial part of the variability (in most cases adjusted R-squared values are 492higher than 0.6), showed that the relationship between income and cancer incidence rate 493remains positive (and significant) even after controlling both for favourable effects of 494economic development - namely population ageing and improvements in cancer detection and 495statistical reporting - and for spatially correlated omitted variables.

496 Another result of this work is that underreporting can be proxied by a concave function 497of physician density, the contribution of which is positive and increasing up to roughly 3.8 498physicians over 1000 inhabitants, while further increases do not seem to be relevant. Only for 499cervix uter cancer did physician density result to be negatively correlated. This should not be 500surprising since prevention is particularly important for cervix uter cancer and physicians 501play a key role in prevention. Finally, omitting to control for physician density would produce 502in some cases inverted-U patterns, with turning points at the very end (or beyond) of the 503income range of the sample. Such an omission, however, was regarded here as a 504misspecification.

505 To sum up, our analysis shows that the cancer epidemic cannot be explained solely by 506higher life expectancy, by better statistics and by regional peculiarities: rather, a significant 507role must also be attributed to environmental degradation and life-styles. Unfortunately, our 508regressions are unable to distinguish between the two. Some clues can be drawn from the case 509of lung cancer, which, despite the decrease in smokers in high-income countries, still exhibits a 510positive relationship with income rather than a Kuznets curve. This could be interpreted as 511arising from environmental exposure. In any case, due to the presence of so many confounding 512factors, separating the environmental effects from the life-style aspects would require either 513using micro data or restricting the analysis to units for which a considerable range of statistics 514is available.

515 However, the relevance of environmental risk factors can be inferred from the 516increasing evidence available from the health literature according to which “environmental 517factors play a more important role in cancer genesis than it is usually agreed” (Irigaray 2007, 518640). Our findings are consistent with this literature, namely, that both social change (e.g.

lifestyles) and "the involuntary exposure to many carcinogens in the environment contributes to the rising trend in cancer incidence" (Belpomme et al. 2007a, 1037).

The policy message that can be drawn from our work is that only by becoming aware of the negative side effects of economic development will we also be able to implement policies to tackle them. This is the message of one of the most important recent reports on cancer, according to which "the realization of just how much disease and ill health can be prevented by focusing on environmental risk factors should add impetus to global efforts to encourage preventive health measures" (Prüss-Üstün, 2016, VII).

ACKNOWLEDGMENTS

We gratefully acknowledge comments and suggestions on earlier versions of this paper received from A. Di Ciaula, D. Fiaschi, C. V. Howard, F. Merletti and colleagues, G. Nebbia, and from participants in the International Society of Ecological Economics (ISEE) 2016 Conference (University of the District of Columbia, Washington DC, June 26-29 2016) and in the "Economics, Economic Policies and Sustainable Growth in the Wake of the Crisis" (Università Politecnica delle Marche, Ancona, Italy, September 8-10 2016). We truly appreciated the work done by the anonymous reviewers who have helped improve the original draft considerably. Errors remain our own.

The present research was partially funded by a PRA 2015 and PRA 2017 grant of the University of Pisa.

The paper is the outcome of the joint work of the authors. Nonetheless, sections 1, 3.1, 3.2, and 5 are attributable to Luzzati, sections 3.2, 3.4 to Parenti, sections 2 and 3.5 to Rughi. The latter also prepared the dataset (including standardisation) and performed a preliminary OLS analysis.

References

Alavanja MC, Samanic C, Dosemeci M, Lubin J, Tarone R, Lynch CF, Knott C, Thomas K, Hoppin JA, Barker J, Coble J, Sandler DP, Blair A, (2003). Use of agricultural pesticides and prostate cancer risk in the Agricultural Health Study cohort. American Journal of Epidemiology, 157(9), 800-814.

Ananthakrishnan AN, Hoffmann RG, & Saeian K (2010). Higher physician density is associated with lower incidence of late-stage colorectal cancer. Journal of general internal medicine, 25(11), 1164-1171.

Anselin L (1988). Spatial Econometrics: Methods and Models. Kluwer: Dordrecht.

Azadeh A and Salibian-Barrera M. (2011). An outlier-robust fit for generalized additive models with applications to disease outbreak detection. Journal of the American Statistical Association 106.494, 719-731.

Baker SG. A cancer theory kerfuffle can lead to new lines of research (2015). Journal of the National Cancer Institute, 107(2), 1-8.

Beaulieu N, Bloom D, Bloom R, and Stein R (2009). Breakaway: The global burden of cancer—challenges and opportunities. The Economist Intelligence Unit.

559Bedessem B, Ruphy S. (2015) SMT or TOFT? How the two main theories of carcinogenesis are
560 made (artificially) incompatible. Acta Biotheor. 63(3):257-67.,

561Belpomme D, Irigaray P, Hardell L, Clapp R, Montagnier L, Epstein S, & Sasco AJ (2007a). The
562 multitude and diversity of environmental carcinogens. Environmental research, 105(3),
563 414-429.

564Belpomme D, Irigaray P, Sasco AJ, Newby JA, Howard V, Clapp R, Hardell L (2007b). The
565 growing incidence of cancer: role of lifestyle and screening detection, Int J Oncol, 30,
566 1037-1049

567Bertram JS (2001). The molecular biology of cancer. Molecular aspects of medicine, 21(6),
568 167-223.

569Bilano V, Gilmour S, Moffiet T, d'Espaignet ET, Stevens GA, Commar A, Tuyl F, Hudson I, and
570 Shibuya K, 2015. Global trends and projections for tobacco use, 1990–2025: an analysis
571 of smoking indicators from the WHO Comprehensive Information Systems for Tobacco
572 Control.*The Lancet*, *385* (9972), pp.966-976.

573Boffetta P (2006). “Human cancer from environmental pollutants: the epidemiological
574 evidence.” Mutation Research/Genetic Toxicology and Environmental Mutagenesis 608.2
575 (2006): 157-162.

576Boyle P, Levin B (2008). World Cancer Report 2008. World Health Organisation.

577Boyle P, and Parkin DM, (1991) "Statistical methods for registries." Cancer registration:
578 principles and methods 95, 126-158.

579Bray F, Ferlay J, Laversanne M, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M,
580 Steliarova−Foucher E, Swaminathan R, Antoni S and Soerjomataram I (2015). Cancer
581 incidence in five continents: inclusion criteria, highlights from volume X and the global
582 status of cancer registration. International Journal of Cancer, 137(9), pp.2060-2071.

583Bray F, Jemal A, Grey N, Ferlay J, and Forman D (2012). Global cancer transitions according to
584 the Human Development Index (2008–2030): a population-based study. The lancet
585 oncology, 13(8), 790-801.

586Burgio E, Migliore L, 2015. Towards a systemic paradigm in carcinogenesis: linking epigenetics
587 and genetics. Molecular biology reports 42.4 pp. 777-790.

588Crutzen, PJ Geology of mankind. Nature 415, no. 6867 (2002), pp. 23-23.

589Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M, & Comparative Risk Assessment
590 collaborating group (Cancers) (2005). Causes of cancer in the world: comparative risk
591 assessment of nine behavioural and environmental risk factors. The Lancet, 366(9499),
592 1784-1793.

593Dinda S (2004). Environmental Kuznets curve hypothesis: a survey. Ecological economics,
594 49(4), 431-455.

595Draper NR, John JA. Influential observations and outliers in regression. Technometrics. 1981
596 Feb 1;23(1):21-6.

597Fallah M and Kharazmi E (2008). Substantial under-estimation in cancer incidence estimates
598 for developing countries due to under-ascertainment in elderly cancer cases. Cancer
599 Letters 2008; 264: 250–255.

600Farahani M, Subramanian SV, and Canning D (2009). The effect of changes in health sector
601 resources on infant mortality in the short-run and the long-run: a longitudinal
602 econometric analysis. Social Science & Medicine, 68(11), 1918-1925.

603Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, and
604 Bray F (2015). Cancer incidence and mortality worldwide: Sources, methods and major
605 patterns in GLOBOCAN 2012. Int. J. Cancer, 136: E359–E386. doi:10.1002/ijc.29210

606Fidler MM, Soerjomataram I, and Bray F (2016). A global view on cancer incidence and
607 national levels of the human development index. Int. J. Cancer, 139: 2436–2446.

608Fleisher JM, Lou JQ, and Farrell M, (2008). Relationship between physician supply and breast
609 cancer survival: a geographic approach. Journal of community health, 33(4), 179-182.

610Friedberg MW, Hussey PS, and Schneider EC (2010). Primary care: a critical review of the
611 evidence on quality and costs of health care. Health Affairs, 29(5), 766-772.

612GLOBOCAN 2012. Estimated cancer incidence, mortality and prevalence worldwide in 2012,
613 WHO, http://globocan.iarc.fr/Default.aspx

614Gore AC, Crews D, Doan LL, La Merrill M, Patisaul H, Zota A (2014). Introduction to endocrine
615 disrupting chemicals (EDCs), Endocrine Society and Ipen, available on line at
616 https://saludsindanio.org/sites/default/files/documents-files/2999/Introduction
617 %20to%20Endocrine%20Disrupting%20Chemicals.pdf

618Greaves M, 2014. An evolutionary foundation for cancer control" in Stewart B and Wild CP
619 (2014)

620Gujarati, DN, 2004. Basic econometrics. The McGraw– Hill.

621Hanahan D, Weinberg RA (2000). "The hallmarks of cancer", Cell 7, 100: 57-70.

622Hanahan D, Weinberg RA (2011). Hallmarks of cancer: the next generation. Cell 4(144):646–
623 674.

624Howard CV, and Staats G (2013). Cancer and Environmental Influences. Encyclopedia of
625 Systems Biology, 172-179.

626Irigaray P, Newby JA, Clapp R, Hardell L, Howard V, Montagnier L, Epstein S, Belpomme D.
627 (2007) Lifestyle-related factors and environmental agents causing cancer: an overview.
628 Biomedicine & Pharmacotherapy. 61(10):640-58.

629LeSage J, and Pace RK (2009). Introduction to Spatial Econometrics. Boca Raton, Florida: CRC
630 Press, Taylor & Francis Group.

631Levshina, N. (2015). How to do linguistics with R: Data exploration and statistical analysis.
632 John Benjamins Publishing Company.

633Li N, Du XL, Reitzel LR, Xu L and Sturgis EM (2013). Impact of enhanced detection on the
634 increase in thyroid cancer incidence in the United States: review of incidence trends by
635 socioeconomic status within the surveillance, epidemiology, and end results registry,
636 1980–2008. Thyroid, 23(1), pp.103-110.

637Li N, Du XL, Reitzel LR, Xu L and Sturgis EM (2013). Impact of enhanced detection on the
638 increase in thyroid cancer incidence in the United States: review of incidence trends by
639 socioeconomic status within the surveillance, epidemiology, and end results registry,
640 1980–2008. Thyroid, 23(1), pp.103-110.

641Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J., Koskenvuo M, Pukkala E,Skytthe A,
642 Hemminki K (2000). Environmental and heritable factors in the causation of cancer—
643 analyses of cohorts of twins from Sweden, Denmark, and Finland. New England Journal
644 of Medicine, 343(2), 78-85.

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, & Darnell J (2000). Molecular Cell Biology WH Freeman. *New York*, available at http://www.ncbi.nlm.nih.gov/books/NBK21662/

Luzzati T, and Orsini M (2009). "Investigating the energy-environmental Kuznets curve." Energy 34.3: 291-300.

Luzzati T (2015). Kuznets Curves, in Wright J.D., International Encyclopedia of the Social & Behavioral Sciences, 2nd edition, Vol 13, 144-149, Elsevier.

Macinko J, Starfield B, and Shi L (2007). Quantifying the health benefits of primary care physician supply in the United States. International Journal of Health Services, 37(1), 111-126.

Mannucci PM, Harari S, Martinelli I, and Franchini M (2015). Effects on health of air pollution: a narrative review. Internal and Emergency Medicine, 10(6), 657-662.

Maule M, and Merletti F (2012). Cancer transition and priorities for cancer control. The Lancet Oncology, 13(8), 745-746.

Mayumi K, and Giampietro M (2010). Dimensions and logarithmic function in economics: A short critical analysis. Ecological Economics, 69(8), 1604-1609.

Mondal MNI, and Shitan M (2014). Relative importance of demographic, socioeconomic and health factors on life expectancy in low-and lower-middle-income countries. Journal of Epidemiology, 24(2), 117-124.

Moynihan R, Doust J, and Henry D (2012). Preventing overdiagnosis: how to stop harming the healthy. Bmj, (e3502).

Newby JA and Vyvyan Howard C. (2005). Environmental influences in cancer aetiology. Journal of Nutritional & Environmental Medicine. 5(2-3), 56-114.

Newby JA, Busby CC, Howard CV, Platt MJ. (2007), The cancer incidence temporality index: an index to show temporal changes in the age of onset of overall and specific cancer (England and Wales, 1971–1999). Biomedicine & Pharmacotherapy, 61(10), 623-30.

Norman RE, Ryan A, Grant K, Sitas F, and Scott JG (2014). Environmental contributors to childhood cancers. J Environ Immunology and Toxicology, 1(4), 12.

Nowell PC (1976). The clonal evolution of tumor cell populations. Science, 194(4260), 23-28.

Omran AR (2005). The epidemiologic transition: a theory of the epidemiology of population change. Milbank Quarterly 83.4: 731-757.

Parkin DM (2006). The evolution of the population-based cancer registry." Nature Reviews Cancer 6.8: 603-612.

Prüss-Üstün A, Wolf J, Cor Prüss alán C, Bos R, Neira M (2016). Preventing disease through healthy environments. World Health Organization, Geneva, Switzerland.

R Development Core Team. (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. ISBN 3-900051-07-0.

Servan-Schreiber D (2008). We can stop the cancer epidemic, The New York Times, Sept. 19, 2008, available at http://www.nytimes.com/2008/09/19/opinion/19iht-edservan.1.16308287.html

Stare SM, & Jozefowicz JJ (2008). The effects of environmental factors on cancer prevalence rates and specific cancer mortality rates in a sample of OECD developed countries. International Journal of Applied Economics, 5(2), 24.

Shi L (2012). The impact of primary care: a focused review. Scientifica, 2012.

Siegel RL, Miller KD, & Jemal A (2016). Cancer statistics, 2016. CA: a cancer journal for clinicians, 66(1), 7-30.

Smil V (2000). Energy in the twentieth century: resources, conversions, costs, uses, and consequences. Annual Review of Energy and the Environment, 25(1), 21-51.

Steffen W, Grinevald J, Crutzen P, and McNeill J (2011). The Anthropocene: conceptual and historical perspectives. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 369(1938), 842-867.

Steliarova-Foucher E, Stiller C, Kaatsch P, Berrino F, Coebergh JW, Lacour B, and Perkin M (2004). Geographical patterns and time trends of cancer incidence and survival among children and adolescents in Europe since the 1970s (the ACCIS project): an epidemiological study. The Lancet, 364(9451), 2097-2105.

Stern DI (2004). The rise and fall of the environmental Kuznets curve. World development, 32(8), 1419-1439.

Stewart B and Wild CP (eds.) (2014). World cancer report 2014. World Health Organisation, Geneva Switzerland.

Sundmacher L and Busse R (2011). The impact of physician supply on avoidable cancer deaths in Germany. A spatial analysis. Health Policy, 103(1), 53-62.

Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, and Jemal A (2015). Global cancer statistics, 2012. CA: A Cancer Journal for Clinicians, 65: 87–108.

Van Alstine J and Neumayer E (2010) “The environmental Kuznets curve”. In: Gallagher KP (ed.) Handbook on trade and the environment, E. Elgar.

Vineis P, and Wild CP (2014). Global cancer patterns: causes and prevention. The Lancet, 383(9916), 549-557.

WHO (2014). Global status report on noncommunicable diseases 2014. World Health Organization, Geneva, Switzerland.

Wood SN (2006). Generalized Additive Models. An Introduction with R. Chapman and Hall, London.

Wood SN (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B), 1(73), 3–36.

722**APPENDIX**
723

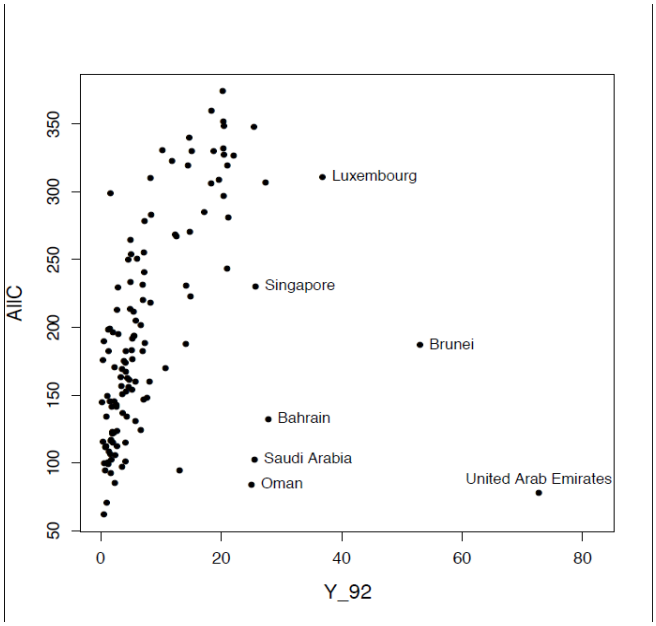
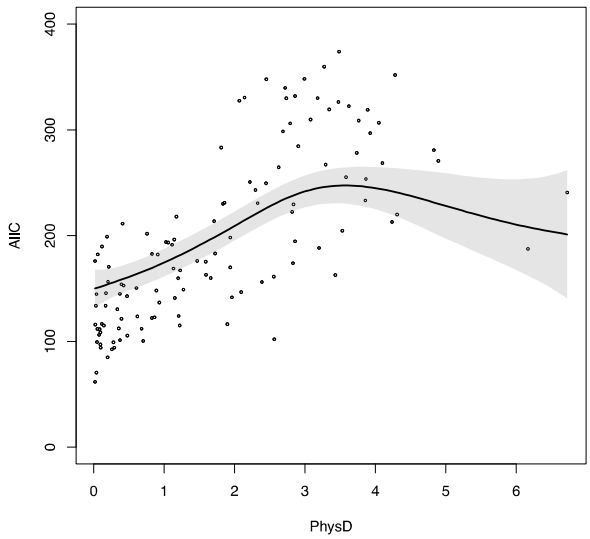


Figure A1. Scatter plot of standardized incidence of all-sites cancers vs. p.c. income in 1992: outliers.



F

igure A2. Marginal impact of Physician density in the semi-parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: all population.

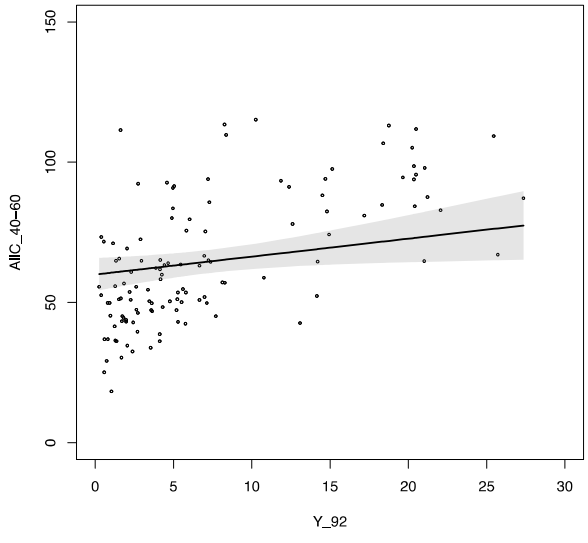


Figure A3. Marginal impact (and 95% confidence band) of p.c. income in the parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: 40-60 yrs.

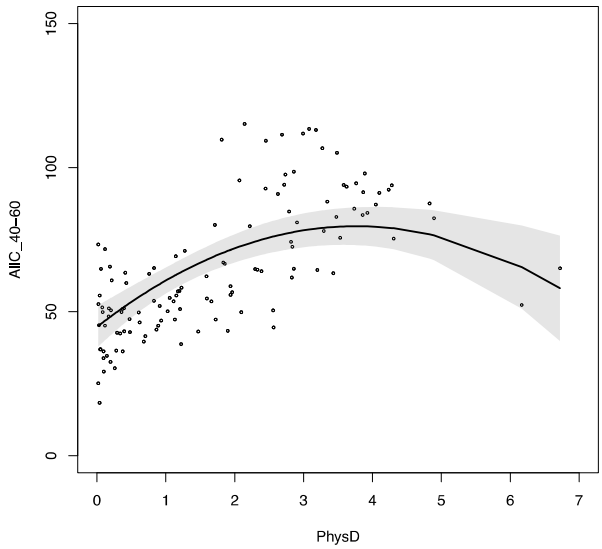


Figure A4. Marginal impact (and 95% confidence band) of physician density in the parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: 40-60 yrs.

724 Table A1. Descriptive statistics of the variables

	Min	Max	Median	Average	Stand Err.
AllC	61.8	374.1	182.2	196.3	78.9
AllC_40_60	18.3	115.2	60.4	64.7	22.7
Lung	0.2	55.5	15.6	17.8	13.2
Lung_40_60	0.1	22.8	4.2	5.1	4.2
Breast	5.4	118.5	45.1	51.4	27.4
Breast_40_60	2.7	57.7	23	25.3	12.8
Colorectum	1.2	48.8	13.1	18.1	13.4
Colorectum_40_60	0.3	16.6	3.9	5.1	3.2
Prostate	1.3	144.4	31.6	44.4	37
Prostate_40_60	0	67.5	4	7.5	9.2
Stomach	0.8	45.4	7.3	9.8	7.6
Stomach_40_60	0.3	15.4	2	2.8	2.4
Liver	1.1	89	5.2	7.7	9.5
Liver_40_60	0.2	29.7	1.4	2.5	3.4
Cervix	2.3	86.7	16.9	20.8	15.4
Cervix_40_60	1.2	49	8.4	9.9	7.7
Oesophagus	0	28	2.6	4.3	4.9
Oesophagus_40_60	0	10.3	0.8	1.3	1.5
Y_92	251	27352	4833	7266	6858
PhysD	0.02	6.72	1.59	1.80	1.50

725 Table A2: Correlation matrix of all-sites incidence rates and regressors

	AllC	A l l C 40-60	*Y_92*	*Y_97*	*Y_02*	*Y_07*	*Y_12*	*Phys*
AllC	1							
AllC 40-60	0.93	1						
Y_92	0.78	0.64	1					
Y_97	0.78	0.63	0.99	1				
Y_02	0.80	0.65	0.98	0.99	1			
Y_07	0.80	0.65	0.96	0.98	0.99	1		
Y_12	0.80	0.66	0.95	0.97	0.98	0.99	1	
Phys	0.71	0.67	0.61	0.58	0.60	0.62	0.63	1

726
727

Table A3. Summary of the Spatial Error Model estimates for the 8 most common organ-sites cancer, 40_60 age classes

1 *Organ Site*	2 *Rel. freq.*	3 *Typical of*	4 *Model*	5 *Role of Y_92*	6 *Role of PhysD*	7 *Y_92 coeff.*	8 *Y_92^2 coeff.*	9 *Adj RSq*
Lung	13.0%	H	Linear	none	+	n.s.		0.61
Breast	11.9%	H	Linear	+	+	0.80***		0.64
Colorectum	9.7%	H	EKC	$	+	0.51***	-0.012**	0.67
Prostate	7.9%	H	Linear	+	n.s.	0.42***		0.47
Stomach	6.8%	M-L	Linear	-	+	-0.17***		0.35
Liver	5.6%	M-L	Linear	-	n.s.	-0.15**		0.11
Cervix uteri	3.7%	M-L	Linear	-	-	-0.26**		0.47
Oesophagus	3.2%	M-L	none	none	n.s.			

Significance p levels: **<0.05, ***<0.01, ^^=0.11; n.s. = non significant
$ calculated turning point = 21,450 $ GDP 1992PPP

742Table A4: Regressions for the most frequent organ-site cancers

743

LUNG

ALL	Coeff.	Std. Error	p
Intercept	5.34	1.86	0.00
Y_92	0.41	0.17	0.01
PhysD	7.40	1.73	0.00
PhysD^2	-0.71	0.30	0.02
	Spatial parameter=0.36*		
	AdjRsq=0.61		
40-60	Coeff.	Std. Error	p
Intercept	2.07	0.72	0.00
Y_92	-0.04	0.06	0.55
PhysD	2.62	0.62	0.00
PhysD^2	-0.26	0.10	0.01
	Spatial parameter=0.47***		
	AdjRsq=0.61		

BREAST

ALL	Coeff.	Std. Error	p
Intercept	22.67	3.87	0.00
Y_92	1.89	0.32	0.00
PhysD	16.23	3.27	0.00
PhysD^2	-2.62	0.55	0.00
	Spatial parameter=0.49***		
	AdjRsq=0.70		
40-60	Coeff.	Std. Error	p
Intercept	11.57	1.89	0.00
Y_92	0.80	0.16	0.00
PhysD	8.93	1.65	0.00
PhysD^2	-1.49	0.28	0.00
	Spatial parameter=0.45***		
	AdjRsq=0.64		

COLORECTUM

ALL	Coeff.	Std. Error	p
Intercept	0.51	1.66	0.76
Y_92	2.25	0.39	0.00
Y_92^2	-0.05	0.02	0.00
PhysD	5.96	1.42	0.00
PhysD^2	-0.84	0.24	0.00
	Spatial parameter=0.39***;		
	AdjRsq=0.76		
40-60	Coeff.	Std. Error	p
Intercept	0.82	0.46	0.07
Y_92	0.51	0.11	0.00
Y_92^2	-0.01	0.00	0.01
PhysD	1.77	0.40	0.00
PhysD^2	-0.27	0.07	0.00
	Spatial parameter=0.36***		
	AdjRsq=0.67		

PROSTATE

ALL	Coeff.	Std. Error	p
Intercept	21.45	6.53	0.00
Y_92	2.43	0.42	0.00
PhysD	4.32	4.24	0.31
PhysD^2	-0.68	0.69	0.32
	Spatial parameter=0.70***		
	AdjRsq=0.75		
40-60	Coeff.	Std. Error	p
Intercept	3.80	1.71	0.03
Y_92	0.42	0.14	0.00
PhysD	0.32	0.59	0.59
	Spatial parameter=0.56***		
	AdjRsq=0.47		

744

745

STOMACH

ALL	Coeff.	Std. Error	p
Intercept	5.37	1.53	0.00
Y_92	-0.37	0.13	0.00
PhysD	6.93	1.32	0.00
PhysD^2	-0.98	0.22	0.00
	Spatial parameter=0.46***		
	AdjRsq=0.36		
40-60	Coeff.	Std. Error	p
Intercept	1.87	0.47	0.00
Y_92	-0.17	0.04	0.00
PhysD	2.15	0.41	0.00
PhysD^2	-0.29	0.07	0.00
	Spatial parameter=0.45***		
	; AdjRsq=0.35		

LIVER

ALL	Coeff.	Std. Error	p
Intercept	8.45	1.80	0.00
Y_92	-0.36	0.18	0.04
PhysD	0.99	0.77	0.20
	Spatial parameter=0.36**		
	AdjRsq=0.08		
40-60	Coeff.	Std. Error	p
Intercept	2.89	0.65	0.00
Y_92	-0.15	0.06	0.01
PhysD	0.38	0.27	0.17
	Spatial parameter=0.38***		
	AdjRsq=0.11		

CERVIX

ALL	Coeff.	Std. Error	p
Intercept	31.11	2.99	0.00
Y_92	-0.39	0.24	0.10
PhysD	-7.03	2.44	0.00
PhysD^2	0.90	0.41	0.03
	Spatial parameter=0.53***		
	AdjRsq=0.48		
40-60	Coeff.	Std. Error	p
Intercept	15.81	1.44	0.00
Y_92	-0.26	0.12	0.03
PhysD	-3.65	1.21	0.00
PhysD^2	0.48	0.20	0.02
	Spatial parameter=0.49***		
	AdjRsq=0.47		

OESOPHAGUS

ALL	Coeff.	Std. Error	p
Intercept	4.51	1.20	0.00
Y_92	-0.21	0.24	0.38
Y_92^2	0.003	0.01	0.77
PhysD	0.50	0.37	0.18
	Spatial parameter=0.57***		
	AdjRsq=0.28		
40-60	Coeff.	Std. Error	p
Intercept	1.49	0.37	0.00
Y_92	-0.08	0.08	0.28
Y_92^2	0.002	0.003	0.58
PhysD	0.10	0.12	0.39
	Spatial parameter=0.54***		

AdjRsq=0.26

List of Countries and Regions

COUNTRIES

Albania	Ecuador	Kenya	Samoa
Algeria	Egypt	Korea, Republic of	Singapore
Armenia	El Salvador	Kyrgyzstan	South African Rep.
Australia	Ethiopia	Lebanon	Spain
Austria	Fiji	Malawi	Sri Lanka
Azerbaijan	Finland	Malaysia	Sudan
Bahamas	France (metrop.).	Mali	Suriname
Bangladesh	FYR Macedonia	Malta	Swaziland
Barbados	Gabon	Mauritius	Sweden
Belarus	Gambia	Mexico	Switzerland
Belgium	Georgia	Moldova, rep. of	Tajikistan
Belize	Germany	Mongolia	Tanzania
Bhutan	Ghana	Morocco	Thailand
Bolivia	Greece	Mozambique	Togo
Botswana	Guatemala	Namibia	Trin. and Tobago
Brazil	Guinea	New Zealand	Tunisia
Bulgaria	Guyana	Nicaragua	Turkey
Burkina Faso	Honduras	Niger	Turkmenistan
Cameroon	Hungary	Nigeria	Uganda
Canada	Iceland	Netherlands, the	Ukraine
Chile	India	Norway	United Kingdom
China	Indonesia	Pakistan	USA
Colombia	Iran, Islamic Rep. of	Panama	Uruguay
Congo, Rep. of	Iraq	Papua New Guinea	Uzbekistan
Costa Rica	Ireland	Paraguay	Vanuatu
Cote d'Ivoire	Israel	Peru	Venezuela
Cuba	Italy	Philippines	Vietnam
Cyprus	Jamaica	Poland	Yemen
Czech Republic	Japan	Portugal	Zambia
Denmark	Jordan	Romania	
Dominican Rep.	Kazakhstan	Russian Federation	

874
875REGIONS
876Australia/New Zealand, Caribbean, Central America, Eastern Africa, Eastern Asia, Eastern
877Europe, Mela/Micro/Polynesia, Middle Africa, Northern Africa, Northern America, Northern
878Europe, South America, South Central Asia, South Eastern Asia, Southern Africa, Southern
879Europe, Western Africa, Western Asia, Western Europe
880