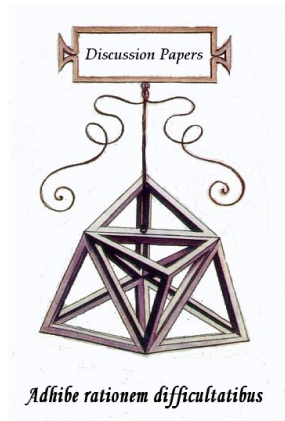

Discussion Papers

Collana di E-papers del Dipartimento di Economia e Management – University of Pisa



Tommaso Luzzati - Angela Parenti – Tommaso Rughi

Spatial error regressions for testing the Cancer-EKC

Discussion Paper n. 218
2017

Discussion Paper n. 218, presentato: July 2017

Indirizzo degli Autori:

Tommaso Luzzati

Dipartimento di Economia e Management, via Ridolfi 10, 56100 PISA – Italy

Email: tommaso.luzzati@unipi.it

Angela Parenti

Institute of Advanced Studies (IMT), Piazza S. Ponziano 6, 55100 LUCCA – Italy

Email: angela.parenti@imtlucca.it

Tommaso Rughi

Dipartimento di Economia e Management, via Ridolfi 10, 56100 PISA – Italy

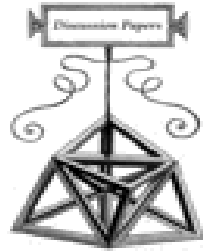
Email: tommaso_rughi@hotmail.it

© *Tommaso Luzzati, Angela Parenti and Tommaso Rughi*

La presente pubblicazione ottempera agli obblighi previsti dall'art. 1 del decreto legislativo luogotenenziale 31 agosto 1945, n. 660.

Please cite as it follows

Tommaso Luzzati, Angela Parenti (2017) and Tommaso Rughi, “Spatial error regressions for testing the Cancer-EKC”, Discussion Papers del Dipartimento di Economia e Management – University of Pisa, n. 218 (<http://www.ec.unipi.it/ricerca/discussion-papers.html>)



Adhuc cursum difficilius

Tommaso Luzzati – Angela Parenti – Tommaso Rughi

Spatial error regressions for testing the Cancer-EKC

Abstract

Why do we observe increasing rates of new cancers cases? Is this mainly the outcome of higher life expectancy and better life conditions brought about by economic development? Do environmental degradation and changes in life-styles play also a role? To answer these questions we empirically assessed the relationship between per capita income and new cancer cases (incidence) by using a cross-sectional dataset from 121 countries.

When looking at the overall incidence rate (i.e., all-sites cancer), we found no support for a cancer-EKC hypothesis (inverted-U relationship). Actually, incidence increases with per capita income, even after controlling for population ageing, improvement in cancer detection, and omitted spatially correlated variables. Hence, a role in cancer occurrence has to be attributed also to changes in lifestyles and to deterioration of environmental quality brought about by economic growth. Looking at the eight most common site-specific cancers not only confirms the existing evidence of different patterns in rich and poor countries, but also helps understanding the estimated relationship for the overall incidence rates.

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

I. Introduction

Cancer incidence (yearly new cases of cancer) is increasing and is predicted to grow at alarming rates, particularly 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).¹ Although data availability² on cancer has increased significantly in the last years³, the relationship between cancer incidence and economic development remains largely unexplored, with just a few exceptions (Beaulieu et al. 2009, Bray et al. 2012, Fidler et al. 2016)⁴. Actually, knowledge about cancer suggests that increase of its incidence is an expected outcome of the increased life expectancy. Moreover, higher levels of development allow also for better health systems, involving improvements both in cancer detections and statistical reporting. At the same time, however, the increase in cancer incidence might be attributed also to “malignant” side effects of economic growth, that is, environmental degradation and bad life-styles.

The aim of the research that is reported here was to better understand, at a macro level, the nature of the relationship between cancer incidence and per capita income. To this purpose, we empirically tested some reduced models that looked only at the ends of complicated causal chains. Such an approach has been followed by the so-called Environmental Kuznets Curve (EKC) literature that, for more than 25 years, has been investigating the relationship between economic growth and the environment (e.g. Stern 2004; Dinda 2004; Luzzati 2015). Differently from the standard EKC approach, rather than using indicators of anthropic pressure, such as emissions or concentrations of pollutants, we focused on a possible direct effect of environmental degradation on human well-being, namely, the cancer effect.

The paper is structured as follows. The second section outlines the possible links between cancer and economic development. The third section describes data and methods. Then, results are presented and discussed. The last section compares our study with similar papers and concludes.

¹For some rich countries incidence rates are stabilizing or slightly decreasing. In the US this has been the case since the mid 1990s (Siegel et al. 2016). Absolute rates are however very high.

²Detailed cancer statistics are available from the Global Cancer Observatory (GCO) interactive web-based platform of the International Agency for Research on Cancer of the World Health Organization, <http://gco.iarc.fr/>. A useful website is also <http://www.wcrf.org/int/cancer-facts-figures/worldwide-data>

³For an assessment of the status of population-based cancer registries worldwide see (Bray et al. 2015).

⁴The differences of the present research with the mentioned ones will be discussed in section 5.

II. 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. A basic idea in 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. Of course, SMT and population ageing explain only part of the story (e.g. Burgio and Migliore). Cancer is increasingly seen as the break of a complex equilibrium, that is, an evolutionary process in which random genetic mutations have to face the selection of environmental pressure; moreover, intrinsic epigenetic plasticity, clonal evolution and high cellular adaptability are also taken into account (Greaves 2016). In other words, cancer is acknowledged as stemming from many complex interacting factors, that is, from mutations in oncogenes and tumour suppressor genes, from genetic inheritance⁵, work and living environment, and lifestyles (see e.g. Belpomme et al. 2007a, Belpomme et al 2007b, Stewart and Wild 2014).

There is general evidence suggesting an increasing role of environmental factors that is estimated to be at around 20% (Prüss-Üstün, p. 16). Although many studies have investigated the differential contribution to cancer incidence of non-genetic risk factors (e.g. Danaei et al. 2005) and of strictly environmental factors (see, e.g., Alavanja 2003, Boffetta 2006, Mannucci et al. 2015, Stare and Jozefowicz 2008), it is very difficult to obtain precise estimates of the role of environmental pollution because of its complexity⁶.

In order to mitigate this difficulty, we used an EKC reduced model where p.c. income is the main regressor. Why should income be a good regressor of cancer incidence? The answer is given by analysing the major links of causal chain that goes from income to cancer, which are described in what follows and summarised in Figure 1. The same figure will prove useful also to illustrate our regression analysis.

⁵The 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 than 42% of the variance in incidence rate, depending on the cancer site.

⁶Environmental pollution is characterized by the multiplicity and variability of polluting agents, the presence of many complex interactions and synergies among agents, the persistence and ubiquity of pollution, the bioaccumulation and bio-magnification of pollutants along the entire food chain, the multiplicity of biological actions performed by each single agent.

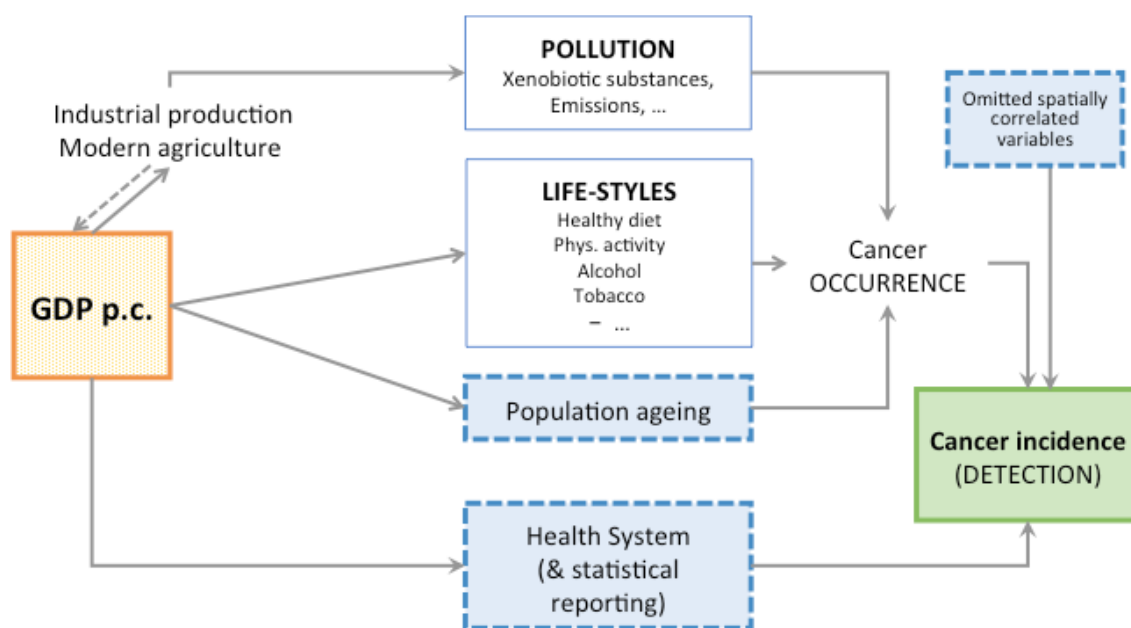


Figure 1: From income to cancer incidence: major links

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, in general, the relationship between humans and nature, to the point that many scholars believe that we entered a new era, 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. New risk factors emerged also due to changes in life-styles occurring along the process of economic development (e.g. excessive-weight and obesity).

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 transition⁷, that is, a shift “from a predominance of

⁷According to the theory of epidemiological transitions (Omran 2005, pp. 737-738), three ages of mortality patterns in history are observed, respectively 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, 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.

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 is important to distinguish between occurrence, detection, and statistical reporting. In countries where health systems are not well developed, both cancer statistics collection is poorly organized, and the causes of death might remain undiagnosed. Since under-registration of cancer deaths is reported in developing countries (Fallah and Kharazmi, 2008), part of the observed increase in cancer incidence could be a result of improved diagnostic scrutiny (e.g., Li et al. 2013, Moynihan et al. 2012).

Figure 1, which summarises the above-discussed links from income to cancer incidence, also represents the idea behind the regression analysis presented in this paper. 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 interpreted as coming, at least partially, from the joint effect of lifestyles and pollution⁸.

More precisely, we regressed cancer incidence rates on the 20 years lagged p.c. income while controlling for (1) population ageing, (2) potential for detection, 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)⁹. In order to control for population ageing, average standardised rates were used to take the different age profiles of the countries into account. Moreover, considering that small size in older age classes in poor countries could cause incidence rates to lose their statistical significance, we did a further check by analysing the age class 40-60 separately. In order to control for improvements in cancer detection and statistical reporting along the process of development, we included physician density (physicians per 1000 inhabitants) in the regressions; the reasons for choosing this variable are discussed in detail in the next section. Many other potential factors (such as genetic risk or diet and habits) can be considered as strongly related to the geographic location of the country. Those factors have

⁸The design of the present analysis does not allow to distinguish lifestyles from environmental conditions. For this purpose, one needs either to use microdata or restrict the analysis to countries for which a rich availability of statistics are available.

⁹Perhaps a longer lag would have been theoretically better, however it would have reduced the number of countries, due to lack of data.

been omitted since they are either unobservable or lacking of reliable data. However a spatial error model was used to take into account these omitted spatially correlated covariates.

III. Material and methods

III.A. Estimation techniques

In the EKC literature different specifications of the p.c. income term (linear, quadratic or cubic) are often tested (Van Alstine and Neumayer, 2010) to choose the one that better fits the data. Less frequently (e.g., Luzzati and Orsini 2009) the specification is chosen by preliminarily running semi-parametric estimates, as we did here. In particular, we used the generalized additive model (GAM) in which each variable enters nonlinearly and separately. The semi parametric estimation was performed by following the approach proposed in Wood (2006), which is based on penalized regression splines. In particular, we used the “mgcv package” in R Development Core Team (2012), with the restricted maximum likelihood (REML) option (see Wood 2011). Once the specification was chosen, we estimated a spatial error regression model to account for potentially omitted covariates that are expected to be spatially correlated (Anselin 1988, LeSage and Pace 2009). A spatial error model was implemented by specifying a spatial stochastic process for the error term, which in turn yields the nonzero correlation for the units that are considered as neighbours. Consequently, the spatial error model requires the definition of a spatial matrix, which reflects the potential interactions between neighbouring units (countries in our case). We used the region classification of the World Health Organization (WHO) to construct the spatial matrix: thus we consider two different countries as neighbours, and therefore interacting with each other, if and only if they belong to the same WHO classification. The spatial regressions are estimated via Maximum Likelihood (Anselin 1988), using the “spdep package” in R Development Core Team (2012).

Variables

Data on cancer incidence are becoming increasingly reliable due to the diffusion of national cancer registries (see, e.g., Parkin and Donald 2009). However, national distinctions in coverage and quality are quite pronounced, resulting in high variability of coverage and reliability of the collected data. For a worldwide comparison, the most relevant project is

GLOBOCAN¹⁰, which is today incorporated in “Cancer today”¹¹. Managed by the WHO’s “International Agency for Research on Cancer”, GLOBOCAN produced the most recent estimates (2012) 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 calculated¹² ASR(W) using the population weight of the World Standard Population¹³ and the population data of the United Nations database.

The p.c. income variable was the p.c. GDP PPP2011¹⁴ of the World Bank online database. Income was averaged over three years to mitigate the effect of the business cycle. As stated in the previous section, we used a 20 years lag to consider the long genesis of cancer. We tested shorter income time lags, which however left results qualitatively unchanged, as one would expect from the strong autocorrelation of p.c. income (see appendix, Table A2).

As regards the variable to proxy the diagnostic potential of a country, a possible candidate would have been p.c. health expenditure. However, the almost perfect linear correlation between p.c. income and health expenditure does not allow the use of both health expenditure and p.c. income as regressors.

According to the literature, physician density has proved to be very relevant to early cancer detection (e.g., Ananthkrishnan et al. 2010, Fleisher et al. 2008, Li et al 2013, Sundmacher and Busse 2011). This holds for many other care issues, such as 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). Of course, early diagnosis increases

¹⁰GLOBOCAN is a project of the International Agency for Research on Cancer of the World Health Organization. Its database contains data for 26 unique site-specific cancers and for all sites cancer (excluding non-melanoma skin cancer). For the scope, the methods, the data sources, and details about the GLOBOCAN project see <http://globocan.iarc.fr/Default.aspx>

¹¹See <http://globocan.iarc.fr/Default.aspx>

¹²The database provided by GLOBOCAN gives already 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 got slightly different figures.

¹³<http://seer.cancer.gov/stdpopulations/world.who.html> World Standard Population is used also in GLOBOCAN 2012.

¹⁴GDP 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).

also with the quality (and expensiveness) of the health system¹⁵. This, however, might be not too relevant in the present analysis since its focus was not early diagnosis but just detection and statistical reporting. For the latter easy access to a doctor is more important than advanced technical tools. At the same time, it is reasonable to assume that physician density has “diminishing returns” in cancer incidence reporting, that is, after some thresholds, further increases in the physician density will have increasingly smaller effects on cancer detection. For these reasons, we took physician density as a proxy of cancer incidence detection potential and used it in the regression analysis with a concave specification. This was supported by the positive and decreasing marginal impact of physician density on cancer incidence in our preliminary semi-parametric estimates (see, e.g., Figure A2).

Physician density (physicians per 1000 inhabitants) was taken from the World Bank online database. Apart from a few exceptions, data for physician density are rather recent, between 2010 and 2013. The correlation (see Table A2) between physician density and GDP p.c. is not strong enough to prevent the use of both variables as regressors.

III.B. 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 five that are not included in the World Bank online database¹⁶. We also excluded 18 countries for which 20 years lagged p.c. income or other data were not available. Finally, seven other countries were considered as outliers since other countries cannot mimic their performances. They have disproportionate p.c. income levels due to very specific economies (based on oil or financial services) and/or their size is very small¹⁷ (see also Figure A1). Hence, our final list, reported in the appendix, included 121 Countries.

III.C. 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

¹⁵Factors such as the presence of screening programmes and diagnostic facilities are relevant. However, for those factors data availability is limited.

¹⁶State of Palestine, France Guadeloupe, France La Reunion, France Martinique, and France Guyana.

¹⁷Bahrain, Brunei, Oman, Saudi Arabia, United Arab Emirates, Luxembourg, and Singapore

the entire population and for the age class 40-60, p.c. income (and its lagged values), and physician density (also as natural logarithm). As expected the autocorrelation of p.c. income is remarkably high.

IV. Results

We start by presenting the results for all cancers, then we move to 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).

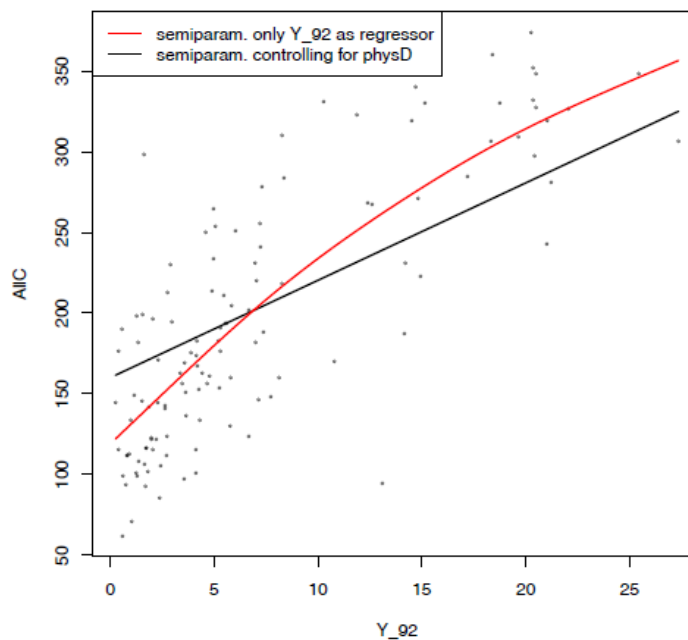


Figure 2. Standardised cancer incidence rates in 2012 vs. p.c. income in 1992. Semiparametric fits when controlling (straight line) and not controlling (curve) for physician density. Age classes: all population.

Figure 2 gives a snapshot of the evidence for all cancers. It shows the scatter plot between the standardised incidence rates for all age classes in 2012 and p.c. income in 1992. A semi parametric regression of the incidence rate on p.c. income gives the concave curve that is shown in Figure 2. However, when controlling for physician density, the marginal impact of p.c. income on cancer incidence becomes linear (the straight line in Figure 2), while the marginal impact of physician density is non-linear (Figure A2). Similar results are obtained when focusing on the 40-60 age class.

This preliminary evidence was helpful to specify the parametric estimates, which, differently from the semiparametric ones, also allow taking into account the possible spatial correlation of errors. Table 1 shows the results of the parametric regressions. For an easier visualization, only the estimates including variables with significant coefficients are presented, implying that the quadratic and cubic terms of income are non significant if not shown.

Table 1. Summary of the Spatial Error Model parametric estimates

Dep. Var.					
<i>AllC</i>	=	106.6	+ 5.3 <i>Y_92</i>	+ 46.6 <i>PhysD</i> - 6.0 <i>PhysD</i> ²	[Eq. 1]
<i>s.e.</i>		9.3	0.9	8.3 1.4	
<i>p.</i>		<0.001	<0.001	<0.001 <0.001	
AdjR ² =0.76; Spatial parameter= 0.42***					
<i>AllC_40-60</i>	=	39.7	+ 0.8 <i>Y_92</i>	+ 18.2 <i>PhysD</i> - 2.4 <i>PhysD</i> ²	[Eq. 2]
<i>s.e.</i>		3.2	0.3	3.0 0.5	
<i>p.</i>		<0.001	<0.001	<0.001 <0.001	
AdjR ² =0.76; Spatial parameter= 0.38***					
<i>AllC</i>	=	130.1	+ 11.5 <i>Y_92</i>	- 0.2 <i>Y_92</i> ²	[Eq. 3]
<i>s.e.</i>		11.7	2.4	0.1	
<i>p.</i>		<0.001	<0.001	0.069	
AdjR ² =0.72; Spatial parameter= 0.55***; calculated turning point <i>Y_92</i> =28.75					
Summary statistics of the observed values:					
			<i>Min</i>	<i>Mean</i>	<i>Max</i>
	<i>AllC</i>	61.8	196.1	374.1	
	<i>AllC_40-60</i>	18.3	64.7	115.2	
	<i>Y_92</i>	0.251	7.113	27.352	
	<i>PhysD</i>	0.02	1.80	6.72	

Equations 1 and 2 in Table 1 show that the incidence rates for all cancers are positively correlated with p.c. income, even after controlling for population ageing, physician density and omitted spatially correlated covariates. Standardised rates increase by 5.3 and by 0.8, respectively for all age and for the age class 40-60, per increase in p.c. income of 1,000\$ (1992, PPP2011). The quadratic (and cubic) terms of income were non significant. Figure 3 and Figure A3 show the marginal impact on standardised incidence rates of p.c. income, drawn respectively for all population and for the 40-60 age class.

The contribution of physician density to fitted incidence rates is measured by the corresponding terms of Equations 1 and 2 and can be visualised by Figure 4 and Figure A4, drawn respectively for all population and for the 40-60 age class. For both regressions, the impact is positive and increasing only up to roughly 3.8 physicians over 1000 inhabitants.

Further increases in physician density beyond this value cannot be assessed since very few countries surpass it and the confidence bands become very wide.

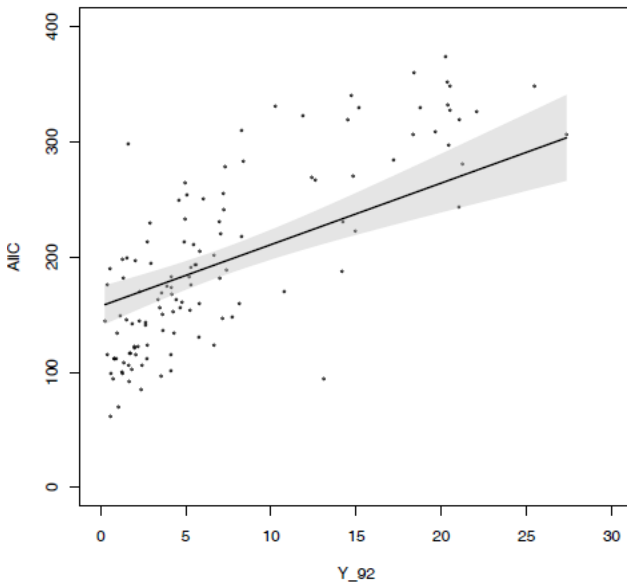


Figure 3. Marginal impact (and 95% confidence band) of p.c. income. (Parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: all population).

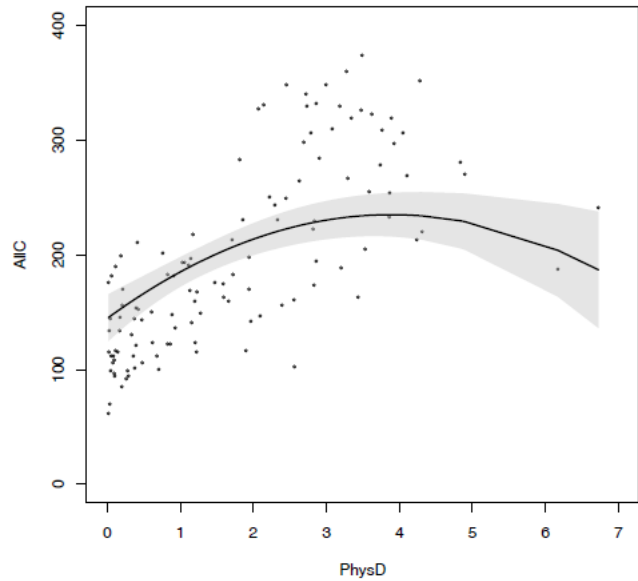


Figure 4. Marginal impact (and 95% confidence band) of physician density. (Parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: all population).

Without controlling for physician density, one would have obtained equation (3). This estimate is consistent with the semiparametric curve reported above in Figure 2. The fitted curve has apparently an inverted-U shape. However, by looking at the observed income values, one notices, that the estimated relationship is increasing within the domain of the observed p.c. income values. Moreover, the reason for the inverted-U shape is that the quadratic term in equation (3) partially captures that which in equation (1) is captured by physician density.

We replicated the analysis illustrated above for the most common site-specific cancers. Table 2 summarises the results. Cancers (second column) are ordered according to their relative frequency, which is shown in the first column. The third column indicates whether the row refers to all age classes or to the 40-60 class. The fourth and fifth columns refer to the type of cancer diffusion, whether typical of low/medium income (L) or high-income countries (H). The fourth column indicates that which is expected on the basis of the health literature, the fifth that which results from our estimates. The sixth column gives a concise indication of the shape of the estimated relationship (and its sign). Notice that the estimated inverted-U relationship for colorectum cancer (EKC) turns down only at the very top of the p.c. income domain. The seventh column shows the estimated coefficients for p.c. income (labelled as 'x').

The eighth column indicates the sign of the coefficient of physician density¹⁸, while the adjusted R-squared is given in the last column. The estimated coefficient of the spatial dependence in the error term is always positive and significant (ranging from 0.34 to 0.68). Accordingly, the error terms were clustered in order to take into account the geographic patterning of measured and unmeasured independent variables.

Our results for the site specific cancers largely not only confirm what is already known about the different pattern of cancers between rich and poor countries, but also give an intuition for the reason why the relationship for all cancers is positive. The sites for which a negative relationship with income holds are relatively less frequent than the sites for which the relationship is positive. When looking at the aggregate incidence, a positive relationship emerges.

Table 2. Summary of the Spatial Error Model estimates for the 8 most common organ-sites cancer

Rel. freq.	Organ Site	Age	Expected	Estimated	Shape	Role of GDP (x)	Phys.	Adj RSq
13.0%	Lung	All	H	H	Linear (+)	$\gamma=5.31+0.41x$	+	0.60
		40-60		No	n.a.	n.s.	+	-
11.9%	Breast	All	L & H	H	Linear (+)	$\gamma=22.2+2.09x$	+	0.70
		40-60		H	Linear (+)	$\gamma=11.51+0.84x$	+	0.63
9.7%	Colorectum	All	H	H	EKC	$\gamma=0.44+2.31x-0.05x^2$	+	0.76
		40-60		H	EKC	$\gamma=0.80+0.52x-0.01x^2$	+	0.67
7.9%	Prostate	All	H	H	Linear (+)	$\gamma=20.82+2.86x$	n.s.	0.75
		40-60		H	Linear (+)	$\gamma=3.42+0.53x$	n.s.	0.48
6.8%	Stomach	All	L (?)	L	Linear (-)	$\gamma=5.33-0.37x$	+	0.36
		40-60		L	Linear (-)	$\gamma=1.85-0.17x$	+	0.35
5.6%	Liver	All	L	L	Linear (-)	$\gamma=8.60-0.41x$	n.s.	0.08
		40-60		L	Linear (-)	$\gamma=2.92-0.17x$	n.s.	0.11
3.7%	Cervix uteri	All	L	L	n.a.	n.s.	-	-
		40-60		L	Linear (-)	$\gamma=15.87-0.28x$	-	0.47
3.2%	Oesophagus	All	L	No	n.a.	n.s.	n.s.	-
		40-60		L	min: $x=22.5$	$\gamma=1.50-0.09x+0.002x^2$	n.s.	0.26

V. Conclusion

The results presented in this paper are related to some evidence available from older data and/or previous studies, the primary goal of which, however, was not to explore the relationship between cancer incidence and income growth. Beaulieu et al. (2009) is a report

¹⁸ Rather interestingly, physician density is negatively correlated with cervix uteri cancer. The increase in physicians not only helps to improve cancer statistics but also is associated with improvements in prevention, which is crucial for reducing the incidence of cervix uteri cancer.

by “The Economist” Intelligence Unit on the health and economic burden of cancer. As a supplementary result, it shows (in the appendix) the outcome of a multiple regression analysis (OLS) aimed at understanding cross-country variations in 2009 estimated cancer incidence rates and in 2002 case fatality rates. Regressors included p.c. income, per cent of population aged 65+, and regional dummies. The paper by Beaulieu et al. (2009) use a methodology similar to ours, since the authors used p.c. income and performed a regression analysis. However, they made use of their own estimates for incidence rates (for year 2009) and controlled for the effect of population ageing by including in the regressions the per cent of population aged 65+.¹⁹

Bray et al. (2012) and Fidler et al. (2016) used the Human Development Index (HDI), which is built by including also life expectancy, rather than p.c. income. Furthermore, they do not use regression analysis; rather they group countries according to the four levels of HDI (low, medium, high, and very high), and compare across groups the incidence and mortality rates. 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 research presented here, because of its methodological differences with the existing literature, not only confirms and updates but also strengthens previous evidence. Our regression analysis showed that the relationship between income and cancer incidence rate remains positive (and significant) even after controlling both for positive effects of economic development, namely population ageing and improvements in cancer detection/statistical reporting, and for spatially correlated omitted variables. Moreover, most of the regressions explain a relevant part of the variability, showing an adjusted R-squared around or higher than 0.7.

The evidence we found can be seen as consistent with the arguments put forward by much theoretical and epidemiological literature, namely that an important role in cancer occurrence has to be attributed also to social changes (e.g. lifestyles) and to deterioration of environmental quality brought about by economic growth.

ACKNOWLEDGMENTS The present research was partially funded by PRA 2015 grant of the University of Pisa

¹⁹The report does not explicitly state whether the regressions in its appendix G were run with age standardised incidence rates.

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.
- Beaulieu N, Bloom D, Bloom R, and Stein R (2009). Breakaway: The global burden of cancer—challenges and opportunities. The Economist Intelligence Unit.
- Belpomme D, Irigaray P, Hardell L, Clapp R, Montagnier L, Epstein S, & Sasco AJ (2007a). The multitude and diversity of environmental carcinogens. *Environmental research*, 105(3), 414-429.
- Belpomme D, Irigaray P, Sasco AJ, Newby JA, Howard V, Clapp R, Hardell L (2007b). The growing incidence of cancer: role of lifestyle and screening detection, *Int J Oncol.*; 30:1037-1049, 2007
- Bertram JS (2001). The molecular biology of cancer. *Molecular aspects of medicine*, 21(6), 167-223.
- Boffetta, P (2006). "Human cancer from environmental pollutants: the epidemiological evidence." *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 608.2 (2006): 157-162.
- Boyle P, Levin B (2008). *World Cancer Report 2008*. World Health Organisation.
- Boyle P, and Parkin DM, (1991) "Statistical methods for registries." *Cancer registration: principles and methods* 95, 126-158.
- Bray F, Ferlay J, Laversanne M, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M, Steliarova-Foucher E, Swaminathan R, Antoni S and Soerjomataram I (2015). Cancer incidence in five continents: inclusion criteria, highlights from volume X and the global status of cancer registration. *International Journal of Cancer*, 137(9), pp.2060-2071.
- Bray F, Jemal A, Grey N, Ferlay J, and Forman D (2012). Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. *The lancet oncology*, 13(8), 790-801.
- Burgio E, Migliore L, 2015, "Towards a systemic paradigm in carcinogenesis: linking epigenetics and genetics." *Molecular biology reports* 42.4 pp. 777-790.
- Crutzen, P.J. "Geology of mankind." *Nature* 415, no. 6867 (2002), pp. 23-23.
- Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M, & Comparative Risk Assessment collaborating group (Cancers) (2005). Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *The Lancet*, 366(9499), 1784-1793.
- Dinda, S. (2004). Environmental Kuznets curve hypothesis: a survey. *Ecological economics*, 49(4), 431-455.

- Fallah M and Kharazmi E (2008). Substantial under-estimation in cancer incidence estimates for developing countries due to under-ascertainment in elderly cancer cases. *Cancer Letters* 2008; 264: 250–255.
- Farahani M, Subramanian SV, and Canning D (2009). The effect of changes in health sector resources on infant mortality in the short-run and the long-run: a longitudinal econometric analysis. *Social Science & Medicine*, 68(11), 1918-1925.
- Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, and Bray F (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer*, 136: E359–E386. doi:10.1002/ijc.29210
- Fidler MM, Soerjomataram I, and Bray F (2016). A global view on cancer incidence and national levels of the human development index. *Int. J. Cancer*, 139: 2436–2446. doi:10.1002/ijc.30382
- Fleisher JM, Lou JQ, and Farrell M, (2008). Relationship between physician supply and breast cancer survival: a geographic approach. *Journal of community health*, 33(4), 179-182.
- Friedberg MW, Hussey PS, and Schneider EC (2010). Primary care: a critical review of the evidence on quality and costs of health care. *Health Affairs*, 29(5), 766-772.
- GLOBOCAN 2012, Estimated cancer incidence, mortality and prevalence worldwide in 2012, WHO, <http://globocan.iarc.fr/Default.aspx>
- Greaves M, 2016. "An evolutionary foundation for cancer control" in International Agency for Research on Cancer." (in ref. 19, 337-342).
- Hanahan D, Weinberg RA (2000). "The hallmarks of cancer", *Cell* 7, 100: 57-70.
- Hanahan D, Weinberg RA (2011). Hallmarks of cancer: the next generation. *Cell* 4(144):646–674.
- LeSage J, and Pace RK (2009). *Introduction to Spatial Econometrics*. Boca Raton, Florida: CRC Press, Taylor & Francis Group.
- Li N, Du XL, Reitzel LR, Xu L and Sturgis EM (2013). Impact of enhanced detection on the increase in thyroid cancer incidence in the United States: review of incidence trends by socioeconomic status within the surveillance, epidemiology, and end results registry, 1980–2008. *Thyroid*, 23(1), pp.103-110.
- Li N, Du XL, Reitzel LR, Xu L and Sturgis EM (2013). Impact of enhanced detection on the increase in thyroid cancer incidence in the United States: review of incidence trends by socioeconomic status within the surveillance, epidemiology, and end results registry, 1980–2008. *Thyroid*, 23(1), pp.103-110.
- Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J., Koskenvuo M, Pukkala E, Skytthe A, Hemminki K (2000). Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal 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, & 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, & 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.
- Mondal, MNI & 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).
- 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, Corvalán 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.
- Stare, S. M., & Jozefowicz, J. J. (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., & 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.
- 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, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J. and Jemal, A. (2015). Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*, 65: 87–108. doi:10.3322/caac.21262

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.

Appendix

Table A1. Descriptive statistics of the variables

	Min	Max	Median	Average	Stand Err.
AllC	61.8	374.1	182.2	196.1	79.2
AllC_40_60	18.3	115.2	59.9	64.7	22.8
Lung	0.2	55.4	15.2	17.2	13.2
Lung_40_60	0.1	22.8	4.2	5.1	4.2
Breast	5.4	118.5	45.1	51.3	27.4
Breast_40_60	2.7	57.7	22.9	25.2	12.8
Colorectum	1.2	48.8	12.9	17.9	13.3
Colorectum_40_60	0.3	16.6	3.9	5.0	3.2
Prostate	1.3	144.4	31.1	44.4	37.1
Prostate_40_60	0	67.5	4.1	7.6	9.2
Stomach	0.8	45.4	7.3	9.7	7.6
Stomach_40_60	0.3	15.4	2.0	2.8	2.4
Liver	1.1	89.1	5.2	7.7	9.5
Liver_40_60	0.2	29.7	1.4	2.5	3.4
Cervix	2.3	86.7	17.1	20.9	15.3
Cervix_40_60	1.2	49.1	8.4	9.9	7.7
Oesophagus	0	27.9	2.6	4.3	4.9
Oesophagus_40_60	0	10.3	0.8	1.3	1.5
Y_92	251	27352	4768	7113	6676
PhysD	0.02	6.72	1.59	1.80	1.50

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

	A l l C							
	AllC	40-60	Y_92	Y_97	Y_02	Y_07	Y_12	Phys
AllC	1							
AllC 40-60	0.93	1						
Y_92	0.8	0.65	1					
Y_97	0.81	0.65	0.99	1				
Y_02	0.82	0.67	0.98	0.99	1			
Y_07	0.84	0.68	0.96	0.97	0.99	1		
Y_12	0.85	0.7	0.95	0.96	0.98	0.99	1	
Phys	0.71	0.67	0.63	0.6	0.62	0.66	0.67	1

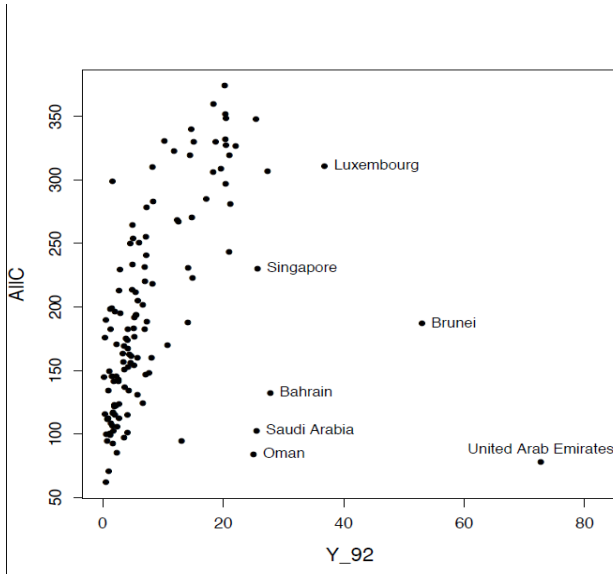


Figure A1. Scatter plot of standardized incidence of all-sites cancers vs. p.c. income in 1992: outliers.

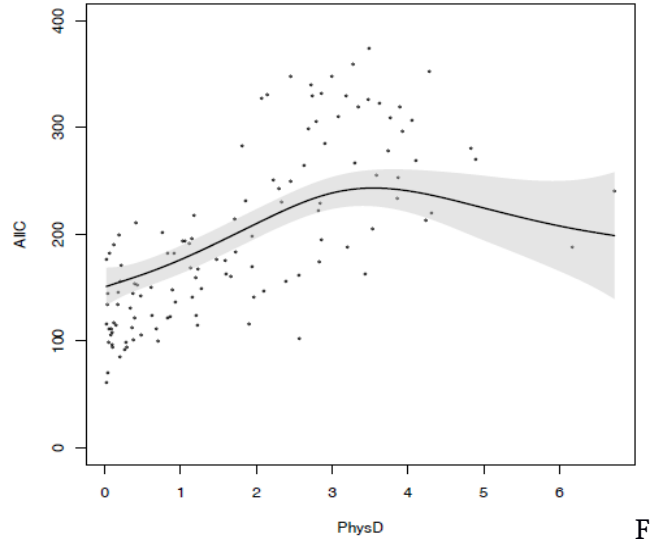


Figure A2. Marginal impact of Physicians density in the semi-parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: all population.

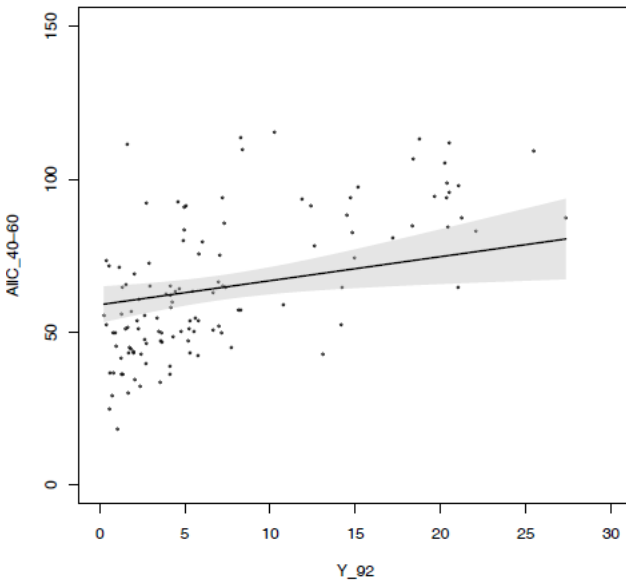


Figure A3. Marginal impact (and 95% confidence band) of p.c. income in the parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: 40-60 yrs.

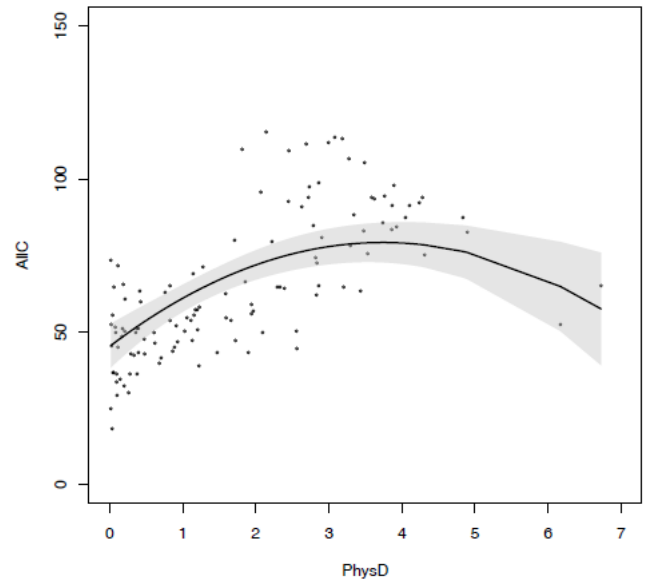


Figure A4. Marginal impact (and 95% confidence band) of physician density in the parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: 40-60 yrs.

List of Countries

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

Discussion Papers

Collana del Dipartimento di Economia e Management, Università di Pisa

Comitato scientifico:

Luciano Fanti - Coordinatore responsabile

Area Economica

Giuseppe Conti
Luciano Fanti
Davide Fiaschi
Paolo Scapparone

Area Aziendale

Mariacristina Bonti
Giuseppe D'Onza
Alessandro Gandolfo
Elisa Giuliani
Enrico Gonnella

Area Matematica e Statistica

Sara Biagini
Laura Carosi
Nicola Salvati

Email della redazione: lfanti@ec.unipi.it