Spatial epidemiology of cancer: a review of data sources, methods and risk factors

Rita Roquette,[1,2] Marco Painho,[1] Baltazar Nunes[2,3]

[1]Nova IMS Information Management School, Lisbon; [2]National Health Institute Doutor Ricardo Jorge, Lisbon; [3]Escola Nacional de Saúde Pública, Universidade Nova de Lisboa, Lisbon, Portugal

Abstract

Cancer is a major concern among chronic diseases today. Spatial epidemiology plays a relevant role in this matter and we present here a review of this subject, including a discussion of the literature in terms of the level of geographic data aggregation, risk factors and methods used to analyse the spatial distribution of patterns and spatial clusters. For this purpose, we performed a websearch in the Pubmed and Web of Science databases including studies published between 1979 and 2015. We found 180 papers from 63 journals and noted that spatial epidemiology of cancer has been addressed with more emphasis during the last decade with research based on data mostly extracted from cancer registries and official mortality statistics. In general, the research questions present in the reviewed papers can be classified into three different sets: i) analysis of spatial distribution of cancer and/or its temporal evolution; ii) risk factors; iii) development of data analysis methods and/or evaluation of results obtained from application of existing methods. This review is expected to help promote research in this area through the identification of relevant knowledge gaps. Cancer's spatial epidemiology represents an important concern, mainly for public health policies design aimed to minimise the impact of chronic disease in specific populations.

Introduction

Given the relevance of spatial epidemiology in health research and the emphasis of cancer among chronic diseases, as well as the growing amount of studies in this area, it is important to know what the literature says about spatial epidemiology of cancer as well as provide its structured description. According to the World Health Organization (WHO), cancer is a leading cause of death in the world (WHO, 2015). It is also the cause of various morbidities and co-morbidities and can be responsible for loss of years of life years as well as loss of years without disability. Considering the aging population, it is predicted that the number of new cases of cancer will increase by more than 12% over the next decade in the European Union (EU) (DGS, 2013). The fight against cancer is a major challenge in public health. This challenge is due in part to the inequalities in terms of incidence, mortality, and survival. Therefore, a multidisciplinary approach is needed (Bastos *et al.*, 2010). Among the various fields that can contribute to the development of knowledge about this disease, spatial epidemiology plays an important role. It can promote the understanding of spatial and temporal distribution patterns, helping to better identify the risk factors that influence them.

Three types of approach can be established in spatial epidemiology: i) mapping; ii) geographic correlation; and iii) clustering (Elliott and Wartenberg, 2004). Mapping or map design regarding health and disease situations is the most often mentioned and used of these three approaches. Further, geographic correlation studies have the goal to spatially compare the health with several types of factors such as environmental, economic, social, demographic or lifestyle (Elliott and Wartenberg, 2004). They can also give clues to the investigation of disease causes (Wakefield, 2004). Finally, concerning the third approach, clustering could be the most relevant from an epidemiologic point of view (Clarke *et al.*, 1996). Cluster can be defined as an *unusual agglomeration* of high or low occurrence of a phenomena (Lawson, 2010).

A search of literature reviews about spatial epidemiology in the Web of Science (Reuteurs, 2016) revealed three main articles, although two of them do not specifically analyse studies related to cancer. Auchincloss and colleagues (2012), in *A Review of Spatial Methods in Epidemiology, 2000-2010*, refer the growing number of articles in the spatial epidemiology field based on articles pub-

lished in seven journals from 2000 to 2010. They also analyse the tools and methods considered in the selected articles. However, they did not specifically analyse cancer-related studies. Further, Boulos and colleagues (2011), in *An eight-year snapshot of geospatial cancer research (2002-2009): clinico-epidemiological and methodological findings and trends*, analyse geospatial cancer research characteristics published in three journals between 2002 and 2009. The analysis focuses on clinical, epidemiological and methodological aspects, namely software used. Finally, Lyssen and colleagues (2014) perform an analysis of the literature about geographic information systems (GIS) and health covering the period from 1991 to 2011 in *A Review and Framework for Categorizing Current Research and Development in Health Related Geographical Information Systems (GIS) Studies*.

This article presents a literature review about cancer's spatial epidemiology. In particular, considering that *literature review* is a generic term, our article presents a *systematized review* (Grant and Booth, 2009). Literature is discussed in terms of the level of geographic data aggregation, risk factors, and methods applied to analyse the spatial distribution of patterns and spatial clusters. The innovation of this study concerns the use of a different approach compared to the reviews described above, *i.e.* it considers cancer specifically and uses a longer period of years of publication to describe the evolution of the volume of published papers and the main subjects covered and also points out gaps of knowledge.

Materials and Methods

We performed a *systematized review* (Grant and Booth, 2009) using the databases Pubmed (accessed at 20th July 2016) and Web of Science (accessed at 28th April 2016). We considered all papers published from 1979 to 2015. Since the search fields available in both databases are different, the search on each one of them was slightly different too. For example, we searched *Title* and *Title/Abstract* in Pubmed and *Title* and *Topic* in the Web of Science. Thus, the search covered all papers published until the end of 2015 that included the following terms: a) ((((((cancer [Title]) OR (neoplasm [Title])) AND epidemiology [Title/Abstract]) AND spati*[Title/Abstract]) AND Geographic* [Title/Abstract]) AND cluster [Title/Abstract]) OR ((((((cancer [Title]) OR (neoplasm [Title])) AND epidemiology

[Title/Abstract]) AND spati*[Title/Abstract]) AND Geographic* [Title/Abstract]) AND distribution [Title/Abstract]) OR ((((((cancer [Title]) OR (neoplasm [Title])) AND epidemiology [Title/Abstract]) AND spati*[Title/Abstract]) AND Geographic* [Title/Abstract]) AND model [Title/Abstract]) OR ((((((cancer[Title]) OR (neoplasm [Title])) AND distribution [Title/Abstract]) AND spati* [Title/Abstract]) AND Geographic* [Title/Abstract]) AND model [Title/Abstract]) in the Pubmed database or; b) ((TI=(cancer) OR TI=(neoplasm)) AND TS=(spati*) AND TS=(epidemiology) AND TS=(geographic*) AND TS=(cluster)) OR ((TI=(cancer) OR TI=(neoplasm)) AND
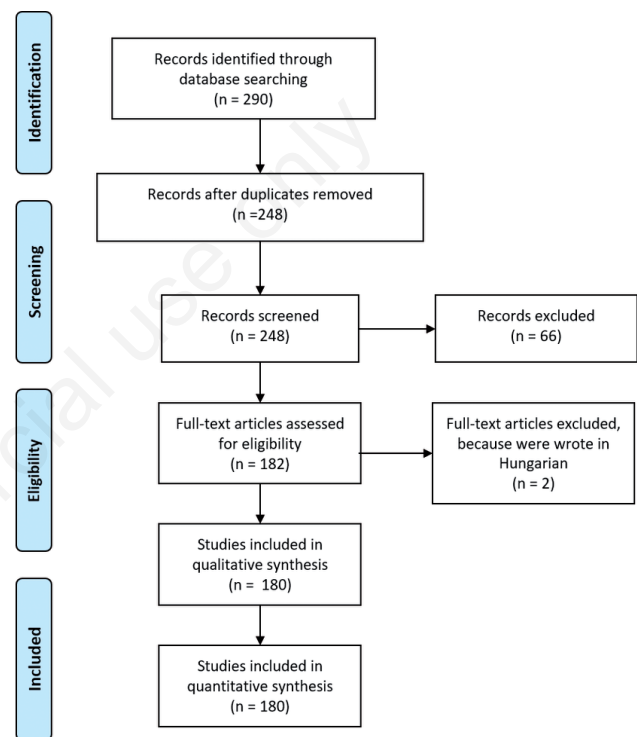

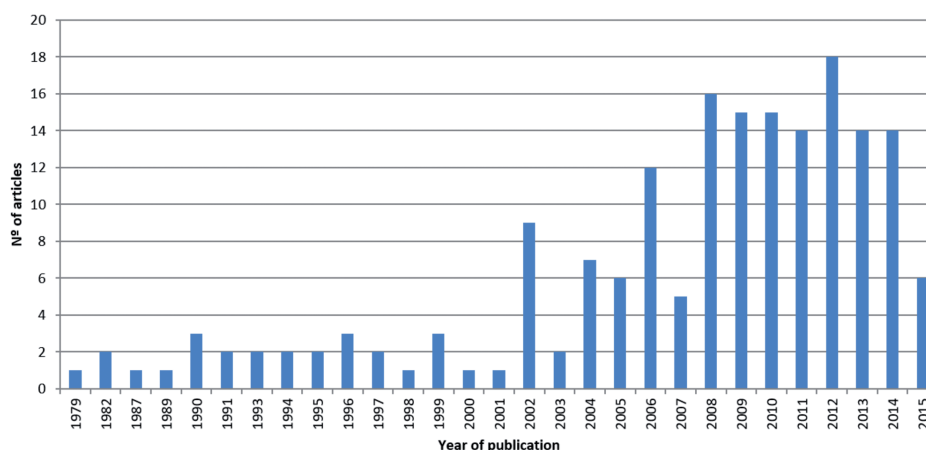**Figure 1. Spatial epidemiology review of cancer: search scheme followed.**

Figure 2. Annual distribution of the published articles selected in the review.

TS=(spati*) AND TS=(epidemiology) AND TS=(geographic*) AND TS=(distribution)) OR ((TI=(cancer) OR TI=(neoplasm)) AND TS=(spati*) AND TS=(epidemiology) AND TS=(geographic*) AND TS=(model)) OR ((TI=(cancer) OR TI=(neoplasm)) AND TS=(spati*) AND TS=(distribution) AND TS=(geographic*) AND TS=(model)), in the Web of Science database.

After article selection, we analysed the scientific areas of the journals in order to refine the search criteria and select the papers more related to the subject addressed. This was performed considering the subject categories associated with each journal at the Scimago website (SCimago, 2007). From the diversity of categories, we selected those considered directly related to the subject of our search, namely Public health, environmental and occupational health; Geography, planning and development; Epidemiology; Oncology; Health (social science); and Health toxicology and mutagenesis. The articles' search scheme, summarising all search steps, is represented in a PRISMA flow diagram (Moher *et al.*, 2010) presented in Figure 1.

The spatial epidemiology of cancer can be very wide-ranging and covered in many different papers. Therefore, as we didn't want to refine our search in terms of publication data, we had to be more restrictive in the choice of search terms. Nevertheless, we tried to ensure that these terms covered a variety of thematic perspectives. The literature was analysed in terms of publication date, keyword, cancer site, data source, observation unit, study objective, risk factor and applied method, as referred in the following sections. The key features of the papers were summarised and described in tables and graphs. Quantitative synthesis was also performed, using descriptive statistics.

Results

General

Our search resulted in a selection of 180 articles from 63 journals. As shown in Figure 2, few articles were published in the early years of the period investigated (1979-2015). The growth in the number of articles started in 2002, grew irregularly until the peak in 2012, with 18 articles published, after which the boom subsided. In terms of the journals with published articles about spatial cancer distribution, Table 1 lists those with more than five articles published.

Key words

Only about 56% of the articles contained keywords. This seems to be related to the journals' publication rules. Although we expect that search options influence the resulting keywords, it is interesting to analyse which ones are the most popular. For this propose, was constructed a *tag cloud* graph [using TagCrowd software (Steinbock, 2006)], which allowed to present the most cited keywords, at a maximum of 50 keywords. The representation of words was made in light of their frequency (Figure 3).

The most frequent keywords were: i) cancer, cited in 14% of the articles; ii) disease mapping, cited in 13%; iii) GIS and epidemiology, each one cited in 12%; iv) breast cancer, spatial analysis and GIS, each one being cited in 11%; v) cancer incidence, lung cancer and prostate cancer, each one presented in 9%; vi) spatial epidemiology in 8%; and, finally, vii) cluster analysis and colorectal cancer, each cited in 6% of the articles. We considered the keywords individually as they were referred to by the authors. However, if the words *geographic information system* and *geographical information system and GIS* were considered together, this keyword becomes the most cited, being present in a total of 26% of the articles.

Table 1. Distribution of articles according to the journal of publication.

Journal	Articles (n)
International Journal of Health Geographics	26
BMC Cancer	10
Cancer Epidemiology	9
Cancer Causes and Control	9
Statistics in Medicine	8
Révue D'Epidemiologie et De Santé Publique	8
Health and Place	6
Environmental Health	6
Spatial and Spatio-Temporal Epidemiology	6

Figure 3. Most cited keywords in the articles selected in the review.

Cancer by site

A total of 28 cancer sites are cited in the reviewed articles. About 59% of the articles refer only one cancer site, and 38% refer two or more sites. The remain articles do not refer to a particular cancer site, considering cancer as a whole, or they indicate various types of cancer but did not mentioned exactly which (Bhowmick *et al.*, 2008; Hendryx *et al.*, 2012; Ruktanonchai *et al.*, 2014). Table 2 summarises the frequency of all cancers analysed, classified according to the 9th revision of International Statistical Classification of Diseases and Related Health Problems (ICD). It provides relative frequencies since an article can address more than one cancer site that makes the interpretation of the number of articles per cancer site difficult.

The cancer sites most frequently studied are malignant neoplasm of bone, connective tissue, skin and breast (especially due breast cancer research) and malignant neoplasm of the genitourinary organs (major part related to prostate cancer analysis).

Data sources used

Data analysis that can be performed on cancer's spatial epidemiology depends first and foremost on the data disaggregation. From the selected articles, six (Bhowmick *et al.*, 2008; Jia *et al.*, 2014; Klassen and Platz, 2006; Lower, 1982; Tuyns and Repetto, 1979; Wan *et al.*, 2012) are essentially theoretical and/or methodological, or did not use cancer data. The others are listed in Table 3 and classified according to data aggregation and data sources.

Most of the articles (about 70%) are based on individual data (Figure 4). Some of these articles used data from both registries and databases from projects or programs (Chien *et al.*, 2013a; Gallagher *et al.*, 2010), while others used individual data and aggregated data (Kulldorff *et al.*, 2006). It should be noticed that a high proportion of the articles used cancer incidence data, *i.e.* all the articles based on aggregated data collected in databases developed within projects or programs and almost all articles representing individual data.

Observation unit used in the articles

As mentioned before, more than half of the articles used data sources of individually disaggregated cancer data. However, in many of these cases, the data were aggregated into areas prior to analysis. Thus, about 74% of the articles have analysed cancer data aggregated by geographic area.

Objectives of the studies presented in the articles

It was found a great variability in the articles objectives. However, independently of their aim or approach, all of them considered somehow geography as an issue. We identified three types of research approaches in the articles (listed and summarised in Table 4). The studies that analysed spatial distribution and/or temporal evolution of disease were applied to concrete and diversified geographical areas. Therefore, we decided not to present a detailed description of their findings. The other two groups of research questions are detailed in the next sections.

Risk factors

The association between cancer morbidity or cancer mortality and possible risk factors is discussed in many of the articles reviewed here (61%), although most of them are not conclusive. The majority provide information regarding the factors that may promote the occurrence of disease but also mention the need of fur-

Table 2. Distribution of articles according to the type of cancer under study indicated as their diagnostic group (ICD 9).

Cancer group (ICD 9)	Percentile
Malignant neoplasm of lip, oral cavity and pharynx	6.7
Malignant neoplasm of digestive organs and peritoneum	18.9
Malignant neoplasm of respiratory and intrathoracic organs	17.3
Malignant neoplasm of bone, connective tissue, skin and breast	22.1
Malignant neoplasm of genitourinary organs	21.8
Malignant neoplasm of other and unspecified sites	6.7
Malignant neoplasm of lymphatic and haematopoietic tissue	6.4

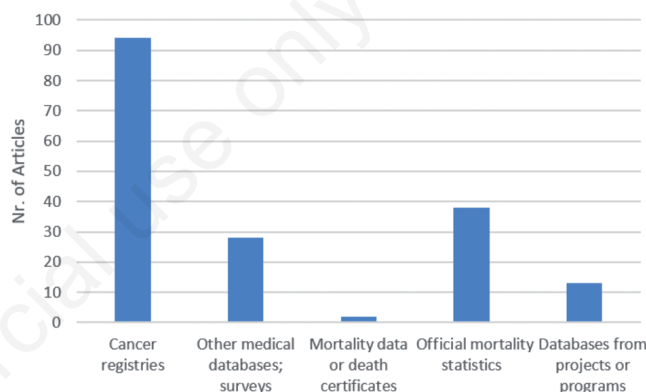ICD, International Classification of Diseases.

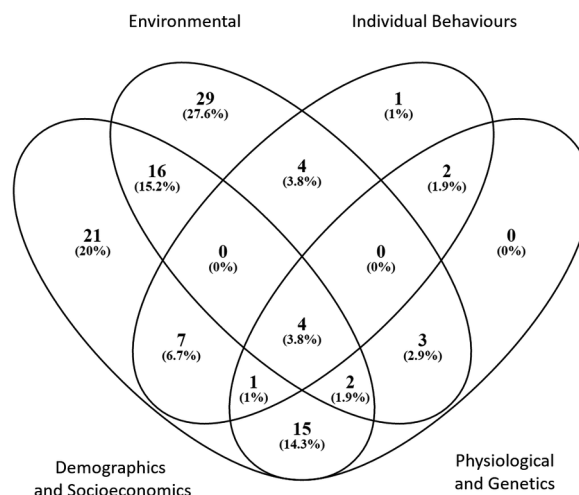**Figure 4. Distribution of articles on cancer morbidity and mortality data source accessed at origin.**

Figure 5. Venn diagram of the number of articles organised by group of cancer risk factor.

ther research to confirm the results. In order to synthesise the factors considered in each article, we classified the factors covered into four groups: demographics and socioeconomics; environmental issues; individual behaviour; and physiological and genetic topics. Results using this classification have been reported in the form of a so called Venn diagram (Oliverus, 2007-2015). Figure 5 shows that demographic and socioeconomic factors together with environmental factors were those most considered for analysis. Further, physiological and genetic factors were analysed more times than individual behaviour. Only very few articles included factors from all the groups.

Table 3. Articles by cancer morbidity and mortality data source accessed at origin.

Data type	Source	Article
Individual data	Cancer registries	Aballay *et al.* (2012), Absalon and Slesak (2011), Al-Ahmadi and Al-Zahrani (2013), Alvarez *et al.* (2009), Bailony *et al.* (2011), Bambhroliya *et al.* (2012), Bristow *et al.* (2014), Buntinx *et al.* (2003), Cassetti *et al.* (2008), Chen *et al.* (2008b, 2011), Chiang *et al.* (2010), Chien *et al.* (2013a), Christian *et al.* (2011), Colak *et al.* (2015), Colonna (2004, 2006), Colonna *et al.* (1990), Cramb *et al.* (2011), Dai and Oyana (2008), David *et al.* (2002), DeChello and Sheehan (2007), DeChello *et al.* (2006), Drapeau *et al.* (1995), Elebead *et al.* (2012), Elferink *et al.* (2012), Fairley *et al.* (2008), Fedewa *et al.* (2009), Ferreira *et al.* (2012), Fortunato *et al.* (2011), Gallagher *et al.* (2010), Garcia Martinez *et al.* (2014), Gbary *et al.* (1995), Goodman *et al.* (2010), Goovaerts and Xiao (2011, 2012), Guajardo and Oyana (2009), Gregorio and Samociuk (2003, 2013), Gregorio *et al.* (2006), Hegarty *et al.* (2010), Hindle *et al.* (2000), Hinrichsen *et al.* (2009), Horner and Chirikos (1987), Horner *et al.* (2011), Hosgood *et al.* (2013), Huang *et al.* (2009), Jarup *et al.* (2002), Johnson (2004), Joseph Sheehan *et al.* (2004), Katayama *et al.* (2014), Knox and Gilman (1996), Kulldorff *et al.* (2006), Luginaah *et al.* (2012), Luo (2013), Luo *et al.* (2010), Sanchez *et al.* (2015), McLafferty and Wang (2009), McLafferty *et al.* (2011), Meliker *et al.* (2009), Mobley *et al.* (2012), Mohebbi *et al.* (2008, 2011), Mueller *et al.* (2015), Nelson (2014), Parodi *et al.* (2005), Pearce *et al.* (2012), Peng *et al.* (2010), Pollack *et al.* (2006), Renart *et al.* (2013), Richardson (1990), Roche *et al.* (2002), Saurina *et al.* (2010), Schuler *et al.* (1991), Sherman *et al.* (2014), Shi (2009), Su *et al.* (2010), Thompson *et al.* (2007, 2008), Thorpe and Shirmohammadi (2005), Tian *et al.* (2012), Ugarte *et al.* (2012, 2015), Voutilainen *et al.* (2014), Walter *et al.* (1994), Xiao *et al.* (2011), Yomralioglu *et al.* (2009), Zhan and Lin (2014), Zhou *et al.* (2008a, 2008b)
	Databases from projects or programmes	Abe *et al.* (2006), Baade *et al.* (2010), Bartolomeo *et al.* (2010), Biggeri *et al.* (1996), Bove *et al.* (2007), Bradshaw *et al.* (1982), Castro *et al.* (2004), Chien *et al.* (2013a), Dasgupta *et al.* (2014), Gallagher *et al.* (2010), Goovaerts (2010), Han *et al.* 2004), Hystad *et al.* (2012, 2013), Knox and Gilman (1996), Krewski *et al.* (2005), Nuemi *et al.* (2013), Ozonoff *et al.* (2005), Paulu *et al.* (2002), Ruktanonchai *et al.* (2014), Scott *et al.* (2002), Shah *et al.* (2014), Sloan *et al.* (2012), Timander and McLafferty (1998), Vieira *et al.* (2005, 2008, 2009), Zlotta *et al.* (2013)
	Death certificates	Chien *et al.* (2013b), Wei *et al.* (2012)
Aggregated data	Official statistics	Aragones *et al.* (2007, 2009, 2013), Berrigan *et al.* (2014), Biggeri *et al.* (2009) Bilancia and Fedespina (2009), Blackley *et al.* 2012), Castro *et al.* (2004), Chen *et al.* (2008a), Chiang *et al.* (2014), Du *et al.* (2010), Fang *et al.* (2004), Godon *et al.* (1991), Goovaerts (2005, 2006a, 2006b), Hendryx *et al.* (2010, 2012), Hosseintabar Marzoni *et al.* (2015), Huang *et al.* (2008), Huang *et al.* (2010), Jacquez and Kheifets (1993), Jemal *et al.* (2002), Kulldorff *et al.* (1997, 2006), Lee and Lin (1994), Lope *et al.* (2008), Lopez-Abente *et al.* (2014), Lopez-Vizcaino *et al.* (2002), Mahoney *et al.* (1990), Pollan *et al.* (2007), Rican *et al.* (1999), Schwartz and Hanchette (2006), Tian *et al.* (2010), Ugarte *et al.* (2015), Vinnakota and Lam (2006), Wojdyla *et al.* (1996), Yang and Hsieh (1998), Zurriaga *et al.* (2008)
	Databases from projects or programmes	Asmarian *et al.* (2013), Boffetta *et al.* (2006), Christakos and Lai (1997), Mahaki *et al.* (2011), Mandal *et al.* (2009), Schootman and Sun (2004), Schootman *et al.* (2011), Senkowski *et al.* (2008), Short *et al.* (2002), Toledano *et al.* (2001), Tsai (2011), Van Leeuwen *et al.* (1999), Yu (2013)

Table 4. Type of research approach used in articles reviewed including summary of themes in each group.

Research approach	Description
Analysis of the spatial distribution and/or temporal evolution of disease	The analysis of geographical distribution patterns and or temporal evolution of disease, covered in about 50% of the articles, can be applied for different purposes, including identification of priority areas for action to fight the disease (Horner *et al.*, 2011), evaluation of the area of influence of health facilities (McLafferty *et al.*, 2011), or, more often, to evaluate the evolution of the number of cases and search for possible explanatory factors for their occurrence (Huang *et al.*, 2009)
Focus on risk factors	This thematic is boarded in about 61% of articles. There are several risk factors of cancer, for example, environmental, socioeconomic, demographic, individual behaviour, or genetic
Development of data analysis methods or evaluation of results of existing applications	Research questions related with thematic of applied analysis methods were present in about 30% of articles, which supports the subject's relevance and the consensus lack about more appropriated methods to the analysis of disease's spatial data

Applied methods in data analysis

All articles that included an analysis of the data presented included also the methodology used. The importance of creating maps that accurately describe disease spatial distribution patterns appeared to be a consensual issue (Kulldorff *et al.*, 2006) though the method used to achieve this was not consensual. Some articles intended to define the best method for some type of analysis for some particular datasets by comparing the results of the application of different spatial analysis methods (Bailony *et al.*, 2011; Biggeri *et al.*, 2009; Chen *et al.*, 2008a; Colonna, 2004; Dasgupta *et al.*, 2014; Goovaerts, 2005, 2006a; Hegarty *et al.*, 2010; Huang *et al.*, 2008; Kaldor and Clayton, 1989; Kulldorff *et al.*, 2006; Meliker *et al.*, 2009; Sherman *et al.*, 2014; Sloan *et al.*, 2012; Zhou *et al.*, 2008b). Table 5 shows a classification of some of most common spatial issues covered by research papers, as well as methods used to answer them.

In the summary presented in Table 5 we did not include any separation by data type (incidence or mortality) or data aggregation (individual data or aggregated) because we could not find any differences in the method applied according to these characteristics. Finally, we need to point out another subject that also seemed to be consensual: the importance of rate standardisation by the individuals' demographic characteristics, particularly by sex and age group. In fact, the standardised rates (by the direct methodology or, more commonly, the indirect one) are used in most articles based on data, regardless of whether they are individual or aggregated. Standardised rates are frequently used in epidemiological studies. The adoption of direct or indirect methods depends mostly on available data. Nevertheless, both allow the comparison between different samples, geographical areas or temporal periods (Bhopal, 2008).

Discussion

The results based on the 180 papers dealing with spatial epidemiology of cancer show that, there was a large increase in the number of papers published in the last decade. This could possibly be due to the increased and now widespread use of computers as well as the generalisation of GIS adoption. GIS appeared in Canada in the 1960s (Tomlinson, 1998) with the aim of acquiring, storing, and processing geographical and alphanumeric information. It allowed the visualisation of both data and results based on such information. GIS usage has become popular in research since the 1990s giving more emphasis to *place* in epidemiological studies (Auchincloss *et al.*, 2012). Nowadays, we can realise that GIS

has a great potential in public health and epidemiology for decision making and research (Clarke *et al.*, 1996).

With regard to the site of the cancers considered in the analysed papers, those most commonly referred to were those known to have the highest incidence rates globally. Some of the most popular keywords that were found in the different papers were: disease mapping; cancer sites (not all sites given but commonly lung and breast); and incidence. This supports the idea that our search criteria seem sufficiently comprehensive to select items with different approaches to the topic under search, using different statistical methods (*e.g.*, mapping, clusters) and addressing several cancer sites with the focus on different epidemiological measures (*e.g.*, incidence).

Cancer incidence data were used in most of the papers reviewed here. The use of incidence rates can be preferable to mortality data from official national statistics since the former can i) provide information on anatomical and histological characteristics of cancer; ii) better describe the extent of the problem of disease in populations; iii) and facilitate comparison of data between countries (Christakos and Lai, 1997). In addition, the survival rate of one cancer site may vary according to geographical area (due to the medical conditions), which may hamper the geographical comparison of mortality data (Horner and Chirikos, 1987).

The analysis of epidemiological study designs could be a very interesting matter in the scope of this study. Nevertheless, a large number of the analysed papers did not clearly describe the epidemiological study design. For that reason, it was not possible to present this information consistently.

As mentioned before, more than a half of the articles reviewed used data sources of individually disaggregated cancer data, which were, in many of the studies, aggregated into areas prior to analysis. Thus, in 74% of the articles, the cancer analysis unit consisted of data aggregated by geographic area. In spatial epidemiology research, the degree of the data's geographical aggregation is a very important issue. Both the use of disaggregated data at the individual level or at large geographical scales (as, for instance when zip codes are used) and the use of aggregated data at small scales have positive and negative aspects. On the one hand, the positive aspects of using disaggregated data are related to the greater variety of possible analytical approaches. There are some analysis methods that are only applicable to individual data (see, for instance, Timander and McLafferty, 1998). On the other, the major problem of using highly disaggregated data is the difficulty of ensuring data confidentiality and anonymity of individuals (Goovaerts, 2005). Also, the use of highly disaggregated geographical data implies a small number of occurrences of a given disease in each area, which makes it difficult to obtain precise statistical

Table 5. Common spatial issues covered by research papers and corresponding methods.

Issue	Method applied
Smoothed rate estimation	Rate calculation in areas with small populations can lead to resultant unstable rates, which can often be mitigated by calculating smoothed rates. The more often referred models in the articles reviewed were Bayesian (both empirical and fully Bayesian models) and Kriging. Among all the models Besag, Mollie and York° stood out
Cluster analysis	In cluster analysis the most used methods were spatial scan statistics, and Moran's *I* (both global and local)
Factor studies	Methods such as hierarchical modelling, multilevel modelling, logistic regression or geographical weighted regression were mentioned in studies of factors or in environmental context analysis

°See, for instance, Lopez-Abente *et al.* (2014) and the webpage: https://www.math.ntnu.no/~hrue/r-inla.org/doc/latent/besag.pdf

values (Chiang *et al.*, 2010; Fairley *et al.*, 2008; Goovaerts, 2006b). This problem, called the *small numbers problem* (Goovaerts, 2005; Shi, 2009), is further enhanced when the diseases investigated are rare (Thompson *et al.*, 2007) and/or the populations of the geographical units under analysis are small (Goovaerts, 2006a; Short *et al.*, 2002).

The difficulties of confidentiality and reliability of highly disaggregated data (Chien *et al.*, 2013b) are more commonly addressed by aggregating the data at small scales or over of long time (several years). One of the benefits of aggregated data at small scales is the mitigation or absence of the *small numbers problem*. However, the larger the data aggregation, the greater the probability of agglomerations with high or low values occupying only part of the area under analysis resulting in hidden information or average dilution of the whole geographic area under investigation (Fang *et al.*, 2004). It is not possible to state, precisely, which degree of data disaggregation is the most appropriate for an analysis in spatial epidemiology and this remains, in fact, a controversial topic. This controversy extends to the question of the stability of calculated statistical measures (like incidence rate). There are studies in which the data are aggregated in order to reduce the uncertainty associated with the analysis results (Huang *et al.*, 2010). However, some authors argue that the spatial pattern of aggregated data could result from aggregation methods rather than data themselves (Krewski *et al.*, 2005).

Regardless the degree of aggregation, the performance of an aggregated data analysis should take into account some concerns, among which the following stand out. First, in combined analysis of geographically aggregated data, difficulties may arise when they are not grouped according to the same geographical boundaries (Blackley *et al.*, 2012; Goovaerts, 2006a); second, analysis results of aggregated data should be considered true only at their scale of aggregation and should not be extrapolated to other aggregation or disaggregation levels (Fortunato *et al.*, 2011) since inconsistencies in results obtained at different scales may arise (Goovaerts and Xiao, 2011, 2012); third, the spatial patterns obtained based on aggregated data can result from the level of aggregation chosen and not from the distribution of the phenomenon under review itself (Krewski *et al.*, 2005); and fourth, data are often aggregated into geographical areas defined for political or administrative reasons (Gregorio *et al.*, 2006), which may not always be the most appropriate for undertaking a particular study (Goovaerts, 2006a). If the areas' aggregation criteria does not take into account the area characteristics in terms of health, the modifiable areal unit problem (MAUP) may arise (Luo, 2013; Shi, 2009; Sloan *et al.*, 2012) and the risk of aggregating areas with very different characteristics could emerge (Thompson *et al.*, 2007).

Thus, the data must be sufficiently disaggregated to allow the researcher to perform the analysis, to obtain statistically robust results, and at the same time not compromising the individual confidentiality (Pearce *et al.*, 2012).

Concerning the spatial risk factors of cancer described in the different papers considered in this review, it turned out to be difficult to identify which factors could promote cancer emergence, since the majority only provide some information and generally emphasise the need for further studies to confirm the results (Jemal *et al.*, 2002). The reasons that could make it difficult to establish a relationship between cancer and spatial risk factors include: i) the latency period of the disease (Jarup *et al.*, 2002; Toledano *et al.*, 2001); ii) the situation in which a factor identified in one geographic region may not have the same effect in another region due to the

presence or absence of other factors (Aragones *et al.*, 2009); or iii) the fact that most cancers result from a combination of several factors rather a single one (Klassen and Platz, 2006).

Among the papers identifying cancer risk factors, we wish to highlight the following: i) the association between arsenic concentration in drinkable water and colon, lung and bladder cancer's incidence risk increase in Cordoba, Argentina (Aballay *et al.*, 2012); ii) a relation between higher incidence and mortality rates by cervical cancer, and more poverty and/or higher distance to screening in USA (Horner *et al.*, 2011); iii) the urban disadvantage in risk of breast, colorectal, lung and prostate cancers in Illinois (McLafferty and Wang, 2009); or iv) a relation between Vitamin D insufficiency and an increase of prostate cancer risk (Schwartz and Hanchette, 2006); v) a possible association among coal mining activities and cancer mortality in West Virginia (Hendryx *et al.*, 2010).

A lack of consensus was actually found in the papers found concerning cancer risk factors, and this was also extended to the methods applied. All articles presented here describe the methods and they vary considerably in many of them. A consensual issue is, however, the importance of creating maps that accurately describe disease spatial distribution patterns. This description can serve as a basis, firstly, for defining the areas in which more detailed studies on the disease aetiology must be carried out (Kulldorff *et al.*, 2006) and, secondly, for identifying areas where interventions are needed to reduce the risk and mitigate the consequences of disease (Klassen and Platz, 2006).

Various methods, such as Bayesian models, Kriging, Spatial Scan Statistics and Moran's *I* have been used in the analysis of spatial distribution patterns of cancer.

Bayesian approaches are sometimes criticised because of their failure to take into account the shape of the geographic areas under analysis. Some authors consider that these methods (Bayesian approaches) should be applied only when the shape of geographic areas are relatively homogeneous (Goovaerts, 2006a). If those areas are heterogeneous, it may be appropriate to use techniques that combine both global and local smoothing to deal with the inherent instability (Colonna, 2004). Other ways to eliminate the effect of this instability can be by applying tests of autocorrelation and spatial heterogeneity at the moment of choosing the Bayesian method (Colonna, 2004) or by using an adaptation of the Poisson Kriging, which includes analysing the size and shape of the geographical areas under study and the population density (Goovaerts, 2006a).

In linear generalised additive models (GAM) the predictor depends on unknown smoothing functions of some predictor variables, shifting the focus on inference about these making them a good alternative, in particular when the analysis includes the individual's residential history (Vieira *et al.*, 2008). The ability to incorporate the individual's residential history is also a strong point noted in Q-statistics (Sloan *et al.*, 2012).

With respect to cluster analysis, some authors consider it advisable to compare the results obtained by applying more than one method or software (Chen *et al.*, 2008a), as this procedure allows a greater degree of certainty that one cluster corresponds to a real aggregation of cases (Bailony *et al.*, 2011). Other authors go further and advise the comparison of results obtained by applying various methods at various scales in order to detect areas of activity in public health (Mohebbi *et al.*, 2008). However, it should be remembered that the choice of the methods to be used depends in the end on the objectives of the study and the type of available

data. More specifically, methods are not only closely related to the data under analysis, but also to their degree of geographic aggregation as well as to the features/factors influencing their distribution. For all of this, the most appropriate method for a given situation may not be the most convenient for another, although the analysis can be similar.

Finaly, although a long search period was considered in this review, it is unlikely to include all published papers related to our subject. The choice of search criteria can have an impact on the final selection of papers, *e.g.*, by requiring the presence of the word *cancer* in the title, which could lead to missing articles of possible relevance for the topic. Therefore, this paper does not exclude the need to be complemented with documents in official websites (*e.g.*, IARC, 2016; NCI, 2016; NIH, 2016) regarding cancer data and research, as well as relevant papers and reports based on them (see, for instance, Ferlay *et al.*, 2015 or Ryerson *et al.*, 2016). Nevertheless, this review may help to promote research in this area, through the identification of some relevant knowledge gaps as well as the description and organisation of the knowledge based on the principal published literature. Moreover, cancer's spatial epidemiology represents a very important concern, mainly for public health policies design aimed to minimise the impact of this chronic disease in specified populations.

Conclusions

Spatial epidemiology of cancer has been addressed in many articles, especially in the last decade, the most common cancer sites being breast, trachea-bronchus-lung, and prostate. Incidence rates were preferred over mortality rates as the epidemiologic frequency measure under study. Although individual data appear predominant in this review (74% of the articles), the units of analysis considered were geographic areas showing aggregation of cases. The research questions considered for analysis belonged to three different sets: i) spatial distribution and/or temporal evolution; ii) cancer risk factors; and iii) applied methods. The spatiotemporal evolution of cancer was covered in 50% of the papers analysis. The most common risk factors studied were demographic, socioeconomic and environmental. The methodological choice depended on data type and the analysis applied; specifically, the methodology was closely related to objectives, the data and their degree of geographic aggregation, including the features/factors influencing their distribution. This literature review comprised a large number of articles published in an extended period of years, which allowed presenting different approaches to spatial issues related to cancer epidemiology. Research on cancer's spatial epidemiology represents a very important issue for decision-making and policies definition to fight one of the most important chronic diseases known.

References

Aballay LR, Diaz MD, Francisca FM, Munoz SE, 2012. Cancer incidence and pattern of arsenic concentration in drinking water wells in Cordoba, Argentina. Int J Environ Heal 22:220-31.

Abe T, Martin IB, Roche LM, 2006. Clusters of census tracts with high proportions of men with distant-stage prostate cancer incidence in New Jersey, 1995 to 1999. Am J Prev Med 30:S60-S6.

Absalon D, Slesak B, 2011. The importance of time of exposure to harmful anthropogenic factors as an element of cancer risk assessment in children. Ecotoxicol Environ Safe 74:967-73.

Al-Ahmadi K, Al-Zahrani A, 2013. NO2 and Cancer Incidence in Saudi Arabia. Int J Env Res Public Health 10:5844-62.

Alvarez PLL, Guerra-Yi ME, Faes C, Alvare YG, Molenberghs G, 2009. Spatial analysis of breast and cervical cancer incidence in small geographical areas in Cuba, 1999-2003. Eur J Cancer Prev 18:395-403.

Aragones N, Goicoa T, Pollan M, Militino AF, Perez-Gomez B, Lopez-Abente G, Ugarte MD, 2013. Spatio-temporal trends in gastric cancer mortality in Spain: 1975-2008. Cancer Epidemiol 37:360-9.

Aragones N, Perez-Gomez B, Pollan M, Ramis R, Vidal E, Lope V, Garcia-Perez J, Boldo E, Lopez-Abente G, 2009. The striking geographical pattern of gastric cancer mortality in Spain: environmental hypotheses revisited. BMC Cancer 9:316.

Aragones N, Ramis R, Pollan M, Perez-Gomez B, Gomez-Barroso D, Lope V, Boldo EI, Garcia-Perez J, Lopez-Abente G, 2007. Oesophageal cancer mortality in Spain: a spatial analysis. BMC Cancer 7:3.

Asmarian NS, Ruzitalab A, Amir K, Masoud S, Mahaki B, 2013. Area-to-Area Poisson Kriging Analysis of Mapping of County-Level Esophageal Cancer Incidence Rates in Iran. Asian Pac J Cancer P 14:11-3.

Auchincloss AH, Gebreab SY, Mair C, Roux AVD, 2012. A review of spatial methods in epidemiology, 2000-2010. Annu Rev Pub Health 33:107-22.

Baade PD, Aitken JF, Ferguson M, Gardiner RA, Chambers SK, 2010. Diagnostic and treatment pathways for men with prostate cancer in Queensland: investigating spatial and demographic inequalities. BMC Cancer 10:452.

Bailony MR, Hararah MK, Salhab AR, Ghannam I, Abdeen Z, Ghannam J, 2011. Cancer registration and healthcare access in West Bank, Palestine: a GIS analysis of childhood cancer, 1998-2007. Int J Cancer 129:1180-9.

Bambhroliya AB, Burau KD, Sexton K, 2012. Spatial analysis of county-level breast cancer mortality in Texas. J Environ Public Health 2012:959343.

Bartolomeo N, Trerotoli P, Serio G, 2010. Lung cancer and COPD rates in Apulia: a multilevel multimember model for smoothing disease mapping. Int J Health Geogr 9:15.

Bastos J, Peleteiro B, Gouveia J, Coleman MP, Lunet N, 2010. The state of the art of cancer control in 30 European countries in 2008. Int J Cancer 126:2700-15.

Berrigan D, Tatalovich Z, Pickle LW, Ewing R, Ballard-Barbash R, 2014. Urban sprawl, obesity, and cancer mortality in the United States: cross-sectional analysis and methodological challenges. Int J Health Geogr 13:3.

Bhopal R, 2008. Concepts of epidemiology: integrating the ideas, theories, principles and methods of epidemiology. Oxford University Press, Oxford, UK.

Bhowmick T, Griffin AL, MacEachren AM, Kluhsman BC, Lengerich EJ, 2008. Informing geospatial toolset design: understanding the process of cancer data exploration and analysis. Health Place 14:576-607.

Biggeri A, Barbone F, Lagazio C, Bovenzi M, Stanta G, 1996. Air pollution and lung cancer in Trieste, Italy: Spatial analysis of risk as a function of distance from sources. Environ Health Persp 104:750-4.

Biggeri A, Catelan D, Dreassi E, 2009. The epidemic of lung cancer in Tuscany (Italy): a joint analysis of male and female mortality

OPEN ACCESS

by birth cohort. Spat Spatiotemporal Epidemiol 1:31-40.

Bilancia M, Fedespina A, 2009. Geographical clustering of lung cancer in the province of Lecce, Italy: 1992-2001. Int J Health Geogr 8:40.

Blackley D, Zheng SM, Ketchum W, 2012. Implementing a weighted spatial smoothing algorithm to identify a lung cancer belt in the United States. Cancer Epidemiol 36:436-8.

Boffetta P, Castaing M, Brennan P, 2006. A geographic correlation study of the incidence of pancreatic and other cancers in Whites. Eur J Epidemiol 21:39-46.

Boulos DNK, Ghali RR, Ibrahim EM, Boulos MNK, AbdelMalik P, 2011. An eight-year snapshot of geospatial cancer research (2002-2009): clinico-epidemiological and methodological findings and trends. Med Oncol 28:1145-62.

Bove GE, Rogerson PA, Vena JE, 2007. Case control study of the geographic variability of exposure to disinfectant byproducts and risk for rectal cancer. Int J Health Geogr 6:18.

Bradshaw E, McGlashan ND, Fitzgerald D, Harington JS, 1982. Analyses of cancer incidence in black gold miners from Southern Africa (1964-79). Brit J Cancer 46:737-48.

Bristow RE, Chang J, Ziogas A, Anton-Culver H, Vieira VM, 2014. Spatial analysis of adherence to treatment guidelines for advanced-stage ovarian cancer and the impact of race and socioeconomic status. Gynecol Oncol 134:60-7.

Buntinx F, Geys H, Lousbergh D, Broeders G, Cloes E, Dhollander D, De Beeck LO, Vanden Brande J, Van Waes A, Molenberghs G, 2003. Geographical differences in cancer incidence in the Belgian province of Limburg. Eur J Cancer 39:2058-72.

Cassetti T, La Rosa F, Rossi L, D'Alo D, Stracci F, 2008. Cancer incidence in men: a cluster analysis of spatial patterns. BMC Cancer 8:344.

Castro M, Vieira V, Assunção R, 2004. Space-time analysis of lung cancer mortality in Southern Brazil. Rev Bras Epidem 7:131-43.

Chen J, Roth RE, Naito AT, Lengerich EJ, Maceachren AM, 2008a. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. Int J Health Geogr 7:57.

Chen Y, Yi Q, Mao Y, 2008b. Cluster of liver cancer and immigration: a geographic analysis of incidence data for Ontario 1998-2002. Int J Health Geogr 7:28.

Cheng EMY, Atkinson PM, Shahani AK, 2011. Elucidating the spatially varying relation between cervical cancer and socio-economic conditions in England. Int J Health Geogr 10:51.

Chiang CT, Hwang YH, Su CC, Tsai KY, Lian IB, Yuan TH, Chang TK, 2010. Elucidating the underlying causes of oral cancer through spatial clustering in high-risk areas of Taiwan with a distinct gender ratio of incidence. Geospat Health 4:231-42.

Chiang CT, Lian IB, Chang YF, Chang TK, 2014. Geospatial Disparities and the Underlying Causes of Major Cancers for Women in Taiwan. Int J Env Res Public Health 11:5613-27.

Chien LC, Schootman M, Pruitt SL, 2013a. The modifying effect of patient location on stage-specific survival following colorectal cancer using geosurvival models. Cancer Cause Control 24:473-84.

Chien LC, Yu HL, Schootman M, 2013b. Efficient mapping and geographic disparities in breast cancer mortality at the county-level by race and age in the U.S. Spat Spatiotemporal Epidemiol 5:27-37.

Christakos G, Lai JJ, 1997. A study of the breast cancer dynamics in North Carolina. Soc Sci Med 45:1503-17.

Christian WJ, Huang B, Rinehart J, Hopenhayn C, 2011. Exploring geographic variation in lung cancer incidence in Kentucky using a spatial scan statistic: elevated risk in the Appalachian coal-mining region. Public Health Rep 126:789-96.

Clarke KC, McLafferty SL, Tempalski BJ, 1996. On epidemiology and geographic information systems: a review and discussion of future directions. Emerg Infect Dis 2:85-92.

Colak EH, Yomralioglu T, Nisanci R, Yildirim V, Duran C, 2015. Geostatistical analysis of the relationship between heavy metals in drinking water and cancer incidence in residential areas in the Black Sea region of Turkey. J Environ Health 77:86-93.

Colonna M, 2004. Bootstrap investigation of the stability of disease mapping of Bayesian cancer relative risk estimations. Eur J Epidemiol 19:761-8.

Colonna M, 2006. Influence of a priori parameters on bayesian relative risks estimations. Spatial distribution of bladder cancer in the urban area of Grenoble. Rev Epidemiol Sante Publique 54:529-42.

Colonna M, Menegoz F, Lutz JM, 1990. Critical analysis of a clustering method applied to the study of geographic distribution of cancer. Rev Epidemiol Sante Publique 38:245-54.

Cramb SM, Mengersen KL, Baade PD, 2011. Identification of area-level influences on regions of high cancer incidence in Queensland, Australia: a classification tree approach. BMC Cancer 11:311.

Dai D, Oyana TJ, 2008. Spatial variations in the incidence of breast cancer and potential risks associated with soil dioxin contamination in Midland, Saginaw, and Bay Counties, Michigan, USA. Environ Health 7:49.

Dasgupta P, Cramb SM, Aitken JF, Turrell G, Baade PD, 2014. Comparing multilevel and Bayesian spatial random effects survival models to assess geographical inequalities in colorectal cancer survival: a case study. Int J Health Geogr 13:36.

David S, Remontet L, Bouvier AM, Faivre J, Colonna M, Esteve J, 2002. How to choose in practice the model of spatial variation of cancer incidence? Examples of digestive cancers from Cote-d'Or "department" - France. Rev Epidemiol Sante Publique 50:413-25.

DeChello LM, Gregorio DI, Samociuk H, 2006. Race-specific geography of prostate cancer incidence. Int J Health Geogr 5:59.

DeChello LM, Sheehan TJ, 2007. Spatial analysis of colorectal cancer incidence and proportion of late-stage in Massachusetts residents: 1995-1998. Int J Health Geogr 6:20.

DGS, 2013. Portugal. Doenças oncológicas em números 2013. Direçao-Geral de Saude, Lisbon, Portugal. Available from: https://www.dgs.pt/estatisticas-de-saude/estatisticas-de-saude/publicacoes/portugal-doencas-oncologicas-em-numeros-2013-pdf.aspx

Drapeau A, Thouez JP, Ghadirian P, 1995. Homogeneity of the spatial-distribution of cancers of the female reproductive-system in Quebec. Soc Sci Med 41:949-55.

Du P, Lemkin A, Kluhsman B, Chen J, Roth RE, MacEachren A, Meyers C, Zurlo JJ, Lengerich EJ, 2010. The roles of social domains, behavioral risk, health care resources, and chlamydia in spatial clusters of US cervical cancer mortality: not all the clusters are the same. Cancer Cause Control 21:1669-83.

Elebead FM, Hamid A, Hilmi HS, Galal H, 2012. Mapping cancer disease using geographical information system (GIS) in Gezira State-Sudan. J Community Health 37:830-9.

Elferink MAG, Pukkala E, Klaase JM, Siesling S, 2012. Spatial variation in stage distribution in colorectal cancer in the Netherlands. Eur J Cancer 48:1119-25.

Elliott P, Wartenberg D, 2004) Spatial epidemiology: current approaches and future challenges. Environ Health Persp 112:998-1006.

Fairley L, Forman D, West R, Manda S, 2008. Spatial variation in prostate cancer survival in the Northern and Yorkshire region of England using Bayesian relative survival smoothing. Brit J Cancer 99:1786-93.

Fang ZX, Kulldorff M, Gregorio DI, 2004. Brain cancer mortality in the United States, 1986 to 1995: a geographic analysis. Neuro Oncol 6:179-87.

Fedewa SA, Soliman AS, Ismail K, Hablas A, Seifeldin IA, Ramadan M, Omar HG, Nriagu J, Wilson ML, 2009. Incidence analyses of bladder cancer in the Nile delta region of Egypt. Cancer Epidemiol 33:176-81.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F, 2015. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136:E359-E86.

Ferreira MAF, Gomes MN, Michels FAS, Dantas AA, Latorre M, 2012. Social inequality in morbidity and mortality from oral and oropharyngeal cancer in the city of Sao Paulo, Brazil: 1997-2008. Cad Saude Publica 28:1663-73.

Fortunato L, Abellan JJ, Beale L, LeFevre S, Richardson S, 2011. Spatio-temporal patterns of bladder cancer incidence in Utah (1973-2004) and their association with the presence of toxic release inventory sites. Int J Health Geogr 10:16.

Gallagher LG, Webster TF, Aschengrau A, Vieira VM, 2010. Using residential history and groundwater modeling to examine drinking water exposure and breast cancer. Environ Health Persp 118:749-55.

Garcia Martinez L, Pollan Santamaria M, Lopez-Abente G, Sanchez Jacob MM, Garcia Palomo A, Gonzalez Martinez R, Honrado Franco E, Martin Sanchez V, 2014. Municipal distribution and trend of the incidence of breast cancer in the health area of Leon, Spain (1996-2010). Rev Esp Salud Publica 88:261-9.

Gbary AR, Philippe P, Ducic S, Beland F, 1995. Spatial distribution of anatomical sites of cancer in a Quebec cancer registry. Soc Sci Med 41:863-72.

Godon D, Lajoie P, Thouez JP, 1991. mortality due to cancers of the brain and lymphatic tissues, and leukemia as a function of agricultural pesticide use in quebec (1976-1985). Can J Public Health 82:174-80.

Goodman JE, Hudson TC, Monteiro RJ, 2010. Cancer cluster investigation in residents near a municipal landfill. Hum Ecol Risk Assess 16:1339-59.

Goovaerts P, 2005. Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. Int J Health Geogr 4:31.

Goovaerts P, 2006a. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. Int J Health Geogr 5:52.

Goovaerts P, 2006b. Geostatistical analysis of disease data: visualization and propagation of spatial uncertainty in cancer mortality risk using Poisson kriging and p-field simulation. Int J Health Geogr 5:7.

Goovaerts P, 2010. Visualizing and testing the impact of place on late-stage breast cancer incidence: A non-parametric geostatistical approach. Health Place 16:321-30.

Goovaerts P, Xiao H, 2011. Geographical, temporal and racial disparities in late-stage prostate cancer incidence across Florida: a

Goovaerts P, Xiao H, 2012. The impact of place and time on the proportion of late-stage diagnosis: the case of prostate cancer in Florida, 1981-2007. Spat Spatiotemporal Epidemiol 3:243-53.

Grant MJ, Booth A, 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 26:91-108.

Gregorio DI, Samociuk H, 2003. Breast cancer surveillance using gridded population units, Connecticut, 1992 to 1995. Ann Epidemiol 13:42-9.

Gregorio DI, Samociuk H, 2013. Prostate cancer incidence in light of the spatial distribution of another screening-detectable cancer. Spat Spatiotemporal Epidemiol 6:1-6.

Gregorio DI, Samociuk H, DeChello L, Swede H, 2006. Effects of study area size on geographic characterizations of health events: prostate cancer incidence in Southern New England, USA, 1994-1998. Int J Health Geogr 5:8.

Guajardo OA, Oyana TJ, 2009. A critical assessment of geographic clusters of breast and lung cancer incidences among residents living near the Tittabawassee and Saginaw Rivers, Michigan, USA. J Environ Public Health 2009:316249.

Han DW, Rogerson PA, Nie J, Bonner MR, Vena JE, Vito D, Muti P, Trevisan M, Edge SB, Freudenheim JL, 2004. Geographic clustering of residence in early life and subsequent risk of breast cancer (United States). Cancer Cause Control 15:921-9.

Hegarty AC, Carsin AE, Comber H, 2010. Geographical analysis of cancer incidence in Ireland: a comparison of two Bayesian spatial models. Cancer Epidemiol 34:373-81.

Hendryx M, Conley J, Fedorko E, Luo JH, Armistead M, 2012. Permitted water pollution discharges and population cancer and non-cancer mortality: toxicity weights and upstream discharge effects in US rural-urban areas. Int J Health Geogr 11:9.

Hendryx M, Fedorko E, Anesetti-Rothermel A, 2010. A geographical information system-based analysis of cancer mortality and population exposure to coal mining activities in West Virginia, United States of America. Geospat Health 4:243-56.

Hindle I, Downer MC, Speight PM, 2000. The temporal and spatial epidemiology of lip cancer in England and Wales. Community Dent Health 17:152-60.

Hinrichsen VL, Klassen AC, Song CH, Kulldorff M, 2009. Evaluation of the performance of tests for spatial randomness on prostate cancer data. Int J Health Geogr 8:41.

Horner MJ, Altekruse SF, Zou ZH, Wideroff L, Katki HA, Stinchcomb DG, 2011. U.S. Geographic Distribution of Prevaccine Era Cervical Cancer Screening, Incidence, Stage, and Mortality. Cancer Epidem Biomar 20:591-9.

Horner RD, Chirikos TN, 1987. survivorship differences in geographical comparisons of cancer mortality - an urban-rural analysis. Int J Epidemiol 16:184-9.

Hosgood HD, 3rd, Farah C, Black CC, Schwenn M, Hock JM, 2013. Spatial and temporal distributions of lung cancer histopathology in the state of Maine. Lung Cancer 82:55-62.

Hosseintabar Marzoni AS, Moghimbeigi A, Faradmal J, 2015. Gastric and esophageal cancers incidence mapping in golestan province, iran: using bayesian-gibbs sampling. Osong Public Health Res Perspect 6:100-5.

Huang B, Dignan M, Han D, Johnson O, 2009. Does Distance Matter? Distance to Mammography Facilities and Stage at Diagnosis of Breast Cancer in Kentucky. J Rural Health 25:366-71.

Huang L, Pickle LW, Das B, 2008. Evaluating spatial methods for investigating global clustering and cluster detection of cancer cases. Stat Med 27:5111-42.

Huang L, Tiwari RC, Pickle LW, Zou Z, 2010. Covariate adjusted weighted normal spatial scan statistics with applications to study geographic clustering of obesity and lung cancer mortality in the United States. Stat Med 29:2410-22.

Hystad P, Demers PA, Johnson KC, Brook J, van Donkelaar A, Lamsal L, Martin R, Brauer M, 2012. Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study. Environ Health 11:22.

Hystad P, Demers PA, Johnson KC, Carpiano RM, Brauer M, 2013. Long-term residential exposure to air pollution and lung cancer risk. Epidemiology 24:762-72.

IARC, 2016. International Agency for Research on Cancer. Available from: https://www.iarc.fr/

Jacquez GM, Kheifets LI, 1993. Synthetic cancer variables and the construction and testing of synthetic risk maps. Stat Med 12:1931-42.

Jarup L, Best N, Toledano MB, Wakefield J, Elliott P, 2002. Geographical epidemiology of prostate cancer in Great Britain. Int J Cancer 97:695-9.

Jemal A, Kulldorff M, Devesa SS, Hayes RB, Fraumeni JF, 2002. A geographic analysis of prostate cancer mortality in the United States, 1970-89. Int J Cancer 101:168-74.

Jia C, James W, Kedia S, 2014. Relationship of racial composition and cancer risks from air toxics exposure in Memphis, Tennessee, U.S.A. Int J Environ Res Public Health 11:7713-24.

Johnson GD, 2004. Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modelling. Int J Health Geogr 3:29.

Joseph Sheehan T, DeChello LM, Kulldorff M, Gregorio DI, Gershman S, Mroszczyk M, 2004. The geographic distribution of breast cancer incidence in Massachusetts 1988 to 1997, adjusted for covariates. Int J Health Geogr 3:17.

Kaldor J, Clayton D, 1989. Role of advanced statistical techniques in cancer mapping. Rec Res Cancer 114:87-98.

Katayama K, Yokoyama K, Yako-Suketomo H, Okamoto N, Tango T, Inaba Y, 2014. Breast cancer clustering in Kanagawa, Japan: a geographic analysis. Asian Pac J Cancer Prev 15:455-60.

Klassen AC, Platz EA, 2006. What can geography tell us about prostate cancer? Am J Prev Med 30:S7-S15.

Knox EG, Gilman EA, 1996. Spatial clustering of childhood cancers in Great Britain. J Epidemiol Commun Health 50:313-9.

Krewski D, Burnett R, Jerrett M, Pope CA, Rainham D, Calle E, Thurston G, Thun M, 2005. Mortality and long-term exposure to ambient air pollution: Ongoing analyses based on the American Cancer Society cohort. J Toxicol Environ Health A 68:1093-109.

Kulldorff M, Feuer EJ, Miller BA, Freedman LS, 1997. Breast cancer clusters in the northeast United States: a geographic analysis. Am J Epidemiol 146:161-70.

Kulldorff M, Song CH, Gregorio D, Samociuk H, DeChello L, 2006. Cancer map patterns - are they random or not? Am J Prev Med 30:37-49.

Lawson AB, 2010. Hotspot detection and clustering: ways and means. Environ Ecol Stat 17:231-45.

Lee WC, Lin RS, 1994. Interactions between birth cohort and urbanization on gastric cancer mortality in Taiwan. Int J Epidemiol 23:252-60.

Lope V, Pollan M, Perez-Gomez B, Aragones N, Vidal E, Gomez-Barroso D, Ramis R, Garcia-Perez J, Cabanes A, Lopez-Abente G, 2008. Municipal distribution of ovarian cancer mortality in Spain. BMC Cancer 8:258.

Lopez-Abente G, Aragones N, Perez-Gomez B, Pollan M, Garcia-Perez J, Ramis R, Fernandez-Navarro P, 2014. Time trends in municipal distribution patterns of cancer mortality in Spain. BMC Cancer 14:535.

Lopez-Vizcaino ME, Vidal-Rodeiro CL, Santiago-Perez MI, Vazquez-Fernandez E, Hervada-Vidal X, 2002. An evaluation of spatio-temporal models for the estimation of the mortality relative risk from breast cancer in Galicia, Spain. J Cancer Epidemiol Prev 7:181-93.

Lower GM, 1982. Concepts in causality: chemically induced human urinary bladder cancer. Cancer 49:1056-66.

Luginaah IN, Gorey KM, Oiamo TH, Tang KX, Holowaty EJ, Hamm C, Wright FC, 2012. A geographical analysis of breast cancer clustering in southern Ontario: generating hypotheses on environmental influences. Int J Environ Health Res 22:232-48.

Luo L, 2013. Impact of spatial aggregation error on the spatial scan analysis: a case study of colorectal cancer. Geospat Health 8:23-35.

Luo L, McLafferty S, Wang FH, 2010. Analyzing spatial aggregation error in statistical models of late-stage cancer risk: a Monte Carlo simulation approach. Int J Health Geogr 9:51.

Lyseen AK, Nohr C, Sorensen EM, Gudes O, Geraghty EM, Shaw NT, Bivona-Tellez C, Group IHGW, 2014. A Review and Framework for Categorizing Current Research and Development in Health Related Geographical Information Systems (GIS) Studies. Yearb Med Inform 9:110-24.

Mahaki B, Mehrabi Y, Kavousi A, Akbari ME, Waldhoer T, Schmid VJ, Yaseri M, 2011. Multivariate disease mapping of seven prevalent cancers in Iran using a shared component model. Asian Pac J Cancer P 12:2353-8.

Mahoney MC, LaBrie DS, Nasca PC, Wolfgang PE, Burnett WS, 1990. Population density and cancer mortality differentials in New York State, 1978-1982. Int J Epidemiol 19:483-90.

Mandal R, St-Hilaire S, Kie JG, Derryberry D, 2009. Spatial trends of breast and prostate cancers in the United States between 2000 and 2005. Int J Health Geogr 8:53.

McLafferty S, Wang F, 2009. Rural reversal? Rural-urban disparities in late-stage cancer risk in Illinois. Cancer 115:2755-64.

McLafferty S, Wang FH, Luo L, Butler J, 2011. Rural-urban inequalities in late-stage breast cancer: spatial and social dimensions of risk and access. Environ Plann B Plann Des 38:726-40.

Meliker JR, Jacquez GM, Goovaerts P, Copeland G, Yassine M, 2009. Spatial cluster analysis of early stage breast cancer: a method for public health practice using cancer registry data. Cancer Cause Control 20:1061-9.

Mobley LR, Kuo TM, Watson L, Brown GG, 2012. Geographic disparities in late-stage cancer diagnosis: multilevel factors and spatial interactions. Health Place 18:978-90.

Mohebbi M, Mahmoodi M, Wolfe R, Nourijelyani K, Mohammad K, Zeraati H, Fotouhi A, 2008. Geographical spread of gastrointestinal tract cancer incidence in the Caspian Sea region of Iran: Spatial analysis of cancer registry data. BMC Cancer 8:137.

Mohebbi M, Wolfe R, Jolley D, Forbes AB, Mahmoodi M, Burton RC, 2011. The spatial distribution of esophageal and gastric cancer in Caspian region of Iran: an ecological analysis of diet and socio-economic influences. Int J Health Geogr 10:13.

Moher D, Liberati A, Tetzlaff J, Altman DG, Grp P, 2010. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 8:336-41.

Mueller GS, Clayton AL, Zahnd WE, Hollenbeck KM, Barrow ME, Jenkins WD, Ruez DR, Jr., 2015. Manuscript title: geospatial analysis of cancer risk and residential proximity to coal mines in Illinois. Ecotoxicol Environ Saf 120:155-62.

NCI, 2016. Survaillence, Epidemiology, and End Results program. Available from: http://seer.cancer.gov/

Nelson EJ, Hughes J, Kulasingam SL, 2014. Spatial patterns of human papillomavirus-associated cancers within the state of Minnesota, 1998-2007. Spat Spatiotemporal Epidemiol 9:13-21.

NIH, 2016. Cancer Data Access System. Available from: https://biometry.nci.nih.gov/cdas/

Nuemi G, Afonso F, Roussot A, Billard L, Cottenet J, Combier E, Diday E, Quantin C, 2013. Classification of hospital pathways in the management of cancer: application to lung cancer in the region of burgundy. Cancer Epidemiol 37:688-96.

Oliverus JC, 2007-2015. Venny. An interactive tool for comparing lists with Venn's diagrams. Available from: http://bioinfogp.cnb.csic.es/tools/venny/index.html

Ozonoff A, Webster T, Vieira V, Weinberg J, Ozonoff D, Aschengrau A, 2005. Cluster detection methods applied to the Upper Cape Cod cancer data. Environ Health 4:19.

Parodi S, Stagnaro E, Casella C, Puppo A, Daminelli E, Fontana V, Valerio F, Vercelli M, 2005. Lung cancer in an urban area in Northern Italy near a coke oven plant. Lung Cancer 47:155-64.

Paulu C, Aschengrau A, Ozonoff D, 2002. Exploring associations between residential location and breast cancer incidence in a case-control study. Environ Health Persp 110:471-8.

Pearce DC, Dowling K, Sim MR, 2012. Cancer incidence and soil arsenic exposure in a historical gold mining area in Victoria, Australia: A geospatial analysis. J Expo Sci Environl Epidemiol 22:248-57.

Peng WX, Chen Y, Jiang QW, Zheng YJ, 2010. Spatial analysis of hepatocellular carcinoma and socioeconomic status in China from a Population-based Cancer Registry. Cancer Epidemiol 34:29-33.

Pollack LA, Gotway CA, Bates JH, Parikh-Patel A, Richards TB, Seeff LC, Hodges H, Kassim S, 2006. Use of the spatial scan statistic to identify geographic variations in late stage colorectal cancer in California (United States). Cancer Cause Control 17:449-57.

Pollan M, Ramis R, Aragones N, Perez-Gomez B, Gomez D, Lope V, Garcia-Perez J, Carrasco JM, Garcia-Mendizabal MJ, Lopez-Abente G, 2007. Municipal distribution of breast cancer mortality among women in Spain. BMC Cancer 7:78.

Renart G, Saez M, Saurina C, Marcos-Gragera R, Ocana-Riola R, Martos C, Barcelo MA, Arribas F, Alcala T, 2013. A common error in the ecological regression of cancer incidence on the deprivation index. Rev Panam Salud Publica 34:83-91.

Reuteurs T, 2016. Web of Science 2015. Thomson Reuters, New York, NY, USA.

Rican S, Salem G, Jougla E, 1999. Spatial variation of lung cancer death in France and spatial scales: importance of changing scale in health geography. Rev Epidemiol Sante Publique 47:411-20.

Richardson S, 1990. A method for testing the significance of geographical correlations with application to industrial lung cancer in France. Stat Med 9:515-28.

Roche LM, Skinner R, Weinstein RB, 2002. Use of a geographic information system to identify and characterize areas with high proportions of distant stage breast cancer. J Public Health Manag Pract 8:26-32.

Ruktanonchai CW, Pindolia DK, Striley CW, Odedina FT, Cottler

LB, 2014. Utilizing spatial statistics to identify cancer hot spots: a surveillance strategy to inform community-engaged outreach efforts. Int J Health Geogr 13:39.

Ryerson AB, Eheman CR, Altekruse SF, Ward JW, Jemal A, Sherman RL, Henley SJ, Holtzman D, Lake A, Noone AM, Anderson RN, Ma JM, Ly KN, Cronin KA, Penberthy L, Kohler BA, 2016. Annual Report to the Nation on the Status of Cancer, 1975-2012, featuring the increasing incidence of liver cancer. Cancer 122:1312-37.

Sanchez VM, Garcia-Sanz M, Garcia-Martinez L, Cabero MD, Campanario-Perez F, Sevilla PG, Iglesias LE, Franco EH, de Rada ACD, Diez FG, Lopez-Abente G, 2015. Hospital incidence, trends and municipal distribution of prostate cancer in health area of Leon, Spain (1996-2010). Rev Esp Salud Publica 89:575-83.

Saurina C, Saez M, Marcos-Gragera R, Barcelo MA, Renart G, Martos C, 2010. Effects of deprivation on the geographical variability of larynx cancer incidence in men, Girona (Spain) 1994-2004. Cancer Epidemiol 34:109-15.

Schootman M, Lian M, Deshpande AD, McQueen A, Pruitt SL, Jeffe DB, 2011. Temporal trends in geographic disparities in small-area-level colorectal cancer incidence and mortality in the United States. Cancer Cause Control 22:1173-81.

Schootman M, Sun D, 2004. Small-area incidence trends in breast cancer. Epidemiology 15:300-7.

Schuler G, Bopp M, Schuler D, 1991. Lung cancer patterns in Switzerland: a search for geographical and occupational patterns. Public Health Rev 19:163-78.

Schwartz GG, Hanchette CL, 2006. UV, latitude, and spatial trends in prostate cancer mortality: all sunlight is not the same (United States). Cancer Cause Control 17:1091-101.

SCimago, 2007. SJR - SCimago Journal and Country Rank. Available from: http://www.scimagojr.com

Scott D, Curtis B, Twumasi FO, 2002. Towards the creation of a health information system for cancer in KwaZulu-Natal, South Africa. Health Place 8:237-49.

Senkowski A, Colonna M, Bicout DJ, 2008. Inframunicipality distribution of lung cancer in the urban area of Grenoble. Rev Epidemiol Sante Publique 56:315-21.

Shah SA, Neoh HM, Rahim SS, Azhar ZI, Hassan MR, Safian N, Jamal R, 2014. Spatial analysis of colorectal cancer cases in Kuala Lumpur. Asian Pac J Cancer Prev 15:1149-54.

Sherman RL, Henry KA, Tannenbaum SL, Feaster DJ, Kobetz E, Lee DJ, 2014. Applying spatial analysis tools in public health: an example using SaTScan to detect geographic targets for colorectal cancer screening interventions. Prev Chronic Dis 11:130264.

Shi X, 2009. A geocomputational process for characterizing the spatial pattern of lung cancer incidence in New Hampshire. Ann Assoc Am Geogr 99:521-33.

Short M, Carlin BP, Bushhouse S, 2002. Using hierarchical spatial models for cancer control planning in Minnesota (United States). Cancer Cause Control 13:903-16.

Sloan CD, Jacquez GM, Gallagher CM, Ward MH, Raaschou-Nielsen O, Nordsborg RB, Meliker JR, 2012. Performance of cancer cluster Q-statistics for case-control residential histories. Spat Spatiotemporal Epidemiol 3:297-310.

Steinbock D, 2006. TagCrowd USA. Available from: tagcrowd.com

Su CC, Lin YY, Chang TK, Chiang CT, Chung JA, Hsu YY, Lian IB, 2010. Incidence of oral cancer in relation to nickel and arsenic concentrations in farm soils of patients' residential areas in Taiwan. BMC Public Health 10:67.

Thompson JA, Carozza SE, Zhu L, 2007. An evaluation of spatial and multivariate covariance among childhood cancer histotypes in Texas (United States). Cancer Cause Control 18:105-13.

Thompson JA, Carozza SE, Zhu L, 2008. Geographic risk modeling of childhood cancer relative to county-level crops, hazardous air pollutants and population density characteristics in Texas. Environ Health 7:45.

Thorpe N, Shirmohammadi A, 2005. Herbicides and nitrates in groundwater of Maryland and childhood cancers: A geographic information systems approach. J Environ Sci Health C 23:261-78.

Tian N, Gaines Wilson J, Benjamin Zhan F, 2010. Female breast cancer mortality clusters within racial groups in the United States. Health Place 16:209-18.

Tian N, Goovaerts P, Zhan FB, Chow TE, Wilson JG, 2012. Identifying risk factors for disparities in breast cancer mortality among African-American and Hispanic women. Womens Health Issues 22:e267-76.

Timander LM, McLafferty S, 1998. Breast cancer in West Islip, NY: a spatial clustering analysis with covariates. Soc Sci Med 46:1623-35.

Toledano MB, Jarup L, Best N, Wakefield J, Elliott P, 2001. Spatial variation and temporal trends of testicular cancer in Great Britain. Brit J Cancer 84:1482-7.

Tomlinson R, 1998. The Canadian Geographic Information System. In: T. Foresman (ed.) The history of geographic information systems: perspectives from the pioneers. Prentice Hall PTR, Upper Saddle River, NJ, USA, pp. 21-32.

Tsai PJ, 2011. Spatial autocorrelation calculations of the nine malignant neoplasms in Taiwan in 2005-2009: a gender comparison study. Chin J Cancer 30:757-65.

Tuyns AJ, Repetto F, 1979. geographic pathology and digestive-tract cancers - present status and prospects. Rev Epidemiol Sante Publique 27:465-77.

Ugarte MD, Adin A, Goicoa T, Lopez-Abente G, 2015. Analyzing the evolution of young people's brain cancer mortality in Spanish provinces. Cancer Epidemiol 39:480-5.

Ugarte MD, Etxeberria J, Goicoa T, Ardanaz E, 2012. Gender-specific spatio-temporal patterns of colorectal cancer incidence in Navarre, Spain (1990-2005). Cancer Epidemiol 36:254-62.

Van Leeuwen JA, Waltner-Toews D, Abernathy T, Smit B, Shoukri M, 1999. Associations between stomach cancer incidence and drinking water contamination with atrazine and nitrate in Ontario (Canada) agroecosystems, 1987-1991. Int J Epidemiol 28:836-40.

Vieira V, Webster T, Weinberg J, Aschengrau A, 2009. Spatial analysis of bladder, kidney, and pancreatic cancer on upper Cape Cod: an application of generalized additive models to case-control data. Environ Health 8:3.

Vieira V, Webster T, Weinberg J, Aschengrau A, Ozonoff D, 2005. Spatial analysis of lung, colorectal, and breast cancer on Cape Cod: an application of generalized additive models to case-control data. Environ Health 4:11.

Vieira VM, Webster TF, Weinberg JM, Aschengrau A, 2008. Spatial-temporal analysis of breast cancer in upper Cape Cod, Massachusetts. Int J Health Geogr 7:46.

Vinnakota S, Lam NS, 2006. Socioeconomic inequality of cancer mortality in the United States: a spatial data mining approach. Int J Health Geogr 5:9.

Voutilainen A, Tolppanen AM, Vehvilainen-Julkunen K, Sherwood PR, 2014. From spatial ecology to spatial epidemiology: modeling spatial distributions of different cancer types with principal coordinates of neighbor matrices. Emerg Themes Epidemiol 11:11.

Wakefield J, 2004. A critique of statistical aspects of ecological studies in spatial epidemiology. Environ Ecol Stat 11:31-54.

Walter SD, Birnie SE, Marrett LD, Taylor SM, Reynolds D, Davies J, Drake JJ, Hayes M, 1994. The geographic variation of cancer incidence in Ontario. Am J Public Health 84:367-76.

Wan N, Zhan FB, Zou B, Chow E, 2012. A relative spatial access assessment approach for analyzing potential spatial access to colorectal cancer services in Texas. Appl Geogr 32(2), 291-299

Wei BG, Jia XJ, Ye BX, Yu JP, Zhang BA, Zhang XW, Lu RA, Dong TR, Yang LS, 2012. Impacts of land use on spatial distribution of mortality rates of cancers caused by naturally occurring asbestos. J Expo Sci Environ Epidemiol 22:516-21.

WHO, 2015. NCD mortality and morbidity. Available from: http://www.who.int/gho/ncd/mortality_morbidity/en/

Wojdyla D, Poletto L, Cuesta C, Badler C, Passamonti ME, 1996. Cluster analysis with constraints: its use with breast cancer mortality rates in Argentina. Stat Med 15:741-6.

Xiao H, Tan F, Goovaerts P, 2011. Racial and geographic disparities in late-stage prostate cancer diagnosis in Florida. J Health Care Poor Unders 22:187-99.

Yang CY, Hsieh YL, 1998. The relationship between population density and cancer mortality in Taiwan. Jpn J Cancer Res 89:355-60.

Yomralioglu T, Colak EH, Aydinoglu AC, 2009. Geo-Relationship between cancer cases and the Environment by GIS: a case study of Trabzon in Turkey. Int J Env Res Public Health 6:3190-204.

Yu B, 2013. Predicting county-level cancer incidence rates and counts in the USA. Stat Med 32:3911-25.

Zhan FB, Lin Y, 2014. Racial/Ethnic, socioeconomic, and geographic disparities of cervical cancer advanced-stage diagnosis in Texas. Womens Health Issues 24:519-27.

Zhou HF, Lawson AB, Hebert JR, Slate EH, Hill EG, 2008a. Joint spatial survival modeling for the age at diagnosis and the vital outcome of prostate cancer. Stat Med 27:3612-28.

Zhou HF, Lawson AB, Hebert JR, Slate EH, Hill EG, 2008b. A Bayesian hierarchical modeling approach for studying the factors affecting the stage at diagnosis of prostate cancer. Stat Med 27:1468-89.

Zlotta AR, Egawa S, Pushkar D, Govorov A, Kimura T, Kido M, Takahashi H, Kuk C, Kovylina M, Aldaoud N, Fleshner N, Finelli A, Klotz L, Sykes J, Lockwood G, van der Kwast TH, 2013. Prevalence of prostate cancer on autopsy: cross-sectional study on unscreened Caucasian and Asian men. J Natl Cancer Inst 105:1050-8.

Zurriaga O, Vanaclocha H, Martinez-Beneito MA, Botella-Rocamora P, 2008. Spatio-temporal evolution of female lung cancer mortality in a region of Spain, is it worth taking migration into account? BMC Cancer 8:35.