Environmental Research 150 (2016) 269–281

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Application of the double kernel density approach to the analysis of cancer incidence in a major metropolitan area

Sarit Davarashvili [a], Marina Zusman [a], Lital Keinan-Boker [b,c], Natalya Rybnikova [a], Zalman Kaufman [b], Barbara G. Silverman [b], Jonathan Dubnov [c], Shai Linn [c], Boris A. Portnov [a,*]

[a] Department of Natural Resources & Environmental Management, Faculty of Management, University of Haifa, Mt. Carmel, Haifa 3498838, Israel
[b] Israel Center for Disease Control, Israel
[c] School of Public Health, University of Haifa, Israel

ARTICLE INFO

Article history:
Received 18 April 2016
Received in revised form
26 May 2016
Accepted 6 June 2016

Keywords:
Lung cancer
Non-Hodgkin lymphoma (NHL)
Standardized Incidence Ratios (SIRs)
Air pollution, industrial proximity
Double Kernel Density (DKD) approach
Greater Haifa Metropolitan Area (GHMA)
Israel

ABSTRACT

Rationale: Although cancer is a main cause of human morbidity worldwide, relatively small numbers of new cancer cases are recorded annually in single urban areas. This makes the association between cancer morbidity and environmental risk factors, such as ambient air pollution, difficult to detect using traditional methods of analysis based on age standardized rates and zonal estimates.
Study goal: The present study investigates the association between air pollution and cancer morbidity in the Greater Haifa Metropolitan Area in Israel by comparing two analytical techniques: the traditional zonal approach and more recently developed Double Kernel Density (DKD) tools. While the first approach uses age adjusted Standardized Incidence Ratios (SIRs) for small census areas, the second approach estimates the areal density of cancer cases, normalized by the areal density of background population in which cancer events occurred. Both analyses control for several potential confounders, including air pollution, proximities to main industrial facilities and socio-demographic attributes.
Results: Air pollution variables and distances to industrial facilities emerged as statistically significant predictors of lung and NHL cancer morbidity in the DKD-based models ($p < 0.05$) but not in the models based on SIRs estimates ($p > 0.2$).
Conclusion: DKD models appear to be a more sensitive tool for assessing potential environmental risks than traditional SIR-based models, because DKD estimates do not depend on *a priory* geographic delineations of statistical zones and produce a smooth and continuous disease 'risk surface' covering the entire study area. We suggest using the DKD method in similar studies of the effect of ambient air pollution on chronic morbidity, especially in cases in which the number of statistical areas available for aggregation and comparison is small and recorded morbidity events are relatively rare.

1. Introduction

Air pollution is known to be associated with cancer morbidity, especially with Lung cancers and possibly also with Non-Hodgkin Lymphomas (NHL): 25–30% of these illnesses worldwide are attributed to different environmental risk factors (Cartwright et al., 1999; Jemal et al., 2011; Norman et al., 2014; Pope, Ezzati and Dockery, 2009; Pruss-Ustun, and Corvalan, 2006; Ramis et al., 2009). Recent WHO report estimates that about 15% of all lung cancers worldwide is attributed to ambient air pollution (Straif et al., 2013).

However, in empirical studies, the association between cancer morbidity and specific environmental risk factors is often difficult to detect. One reason is that a common approach to the investigation of environmental effects on cancer incidence is based on Standardized Incidence Ratios (SIRs), which quantify health outcomes for specific localities, regions or census tracts (Boyle and Parkin, 1991; Curtin and Klein, 1995; Rothman, et al., 2008). However, this approach has weaknesses. When the events of interest are geographically localized and the number of areal units available for aggregation is small, it may lead to a low statistical power and often inconclusive results (Curtin and Klein, 1995; Hollenbeck et al., 2006; Lan and Lian, 2010). In addition, areal aggregation of individual observations may result in information loss and bias due to different weight gains for

* Corresponding author.
E-mail address: portnov@research.haifa.ac.il (B.A. Portnov).

http://dx.doi.org/10.1016/j.envres.2016.06.010
0013-9351/© 2016 Published by Elsevier Inc.

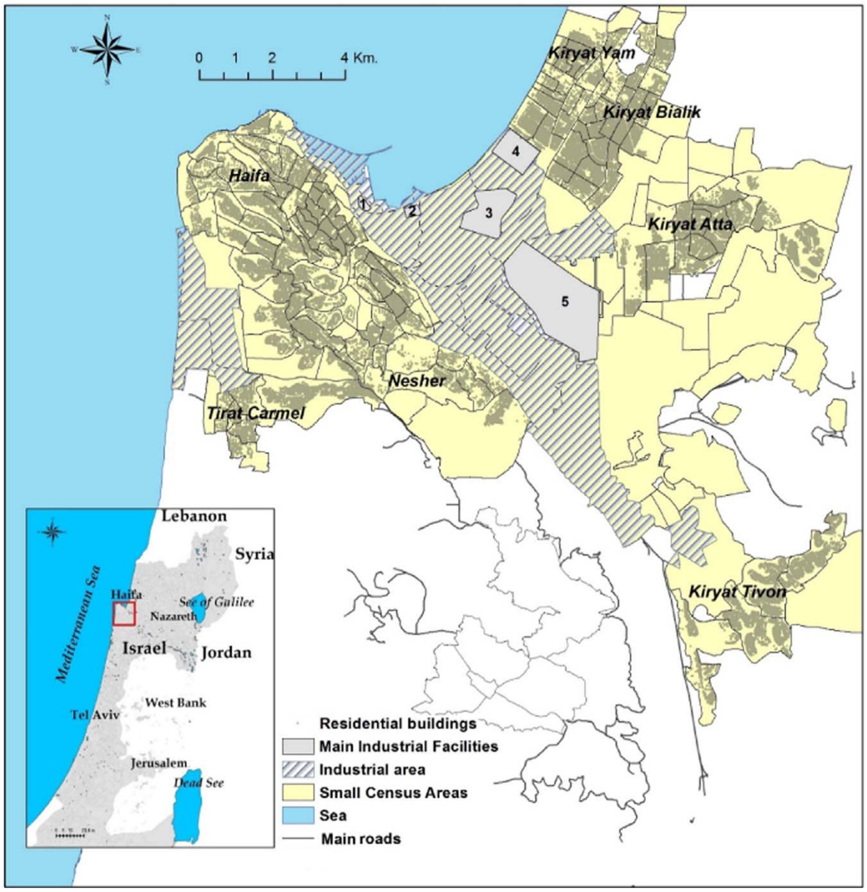

Fig. 1. Map of the study area showing the location of main industrial facilities (1 through 5), small census areas (SCAs), and townships under study.

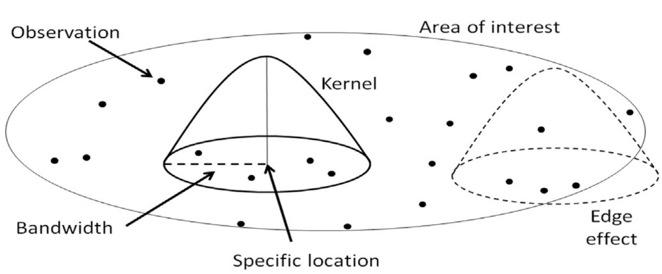

Fig. 2. KD estimation for a point pattern (after Zusman et al. (2015)).

different ‘standard populations’ (Schoenbach and Rosamond, 2000).

The way in which geographic boundaries of statistical areas are delineated may also affect the results of the analysis, reflecting the phenomenon known as the ‘modifiable areal unit problem’ or MAUP (Kloog et al., 2009). The MAUP implies that if the same set of data is aggregated into several sets of areal units, each combination may lead to different inferences, especially if the units of aggregations are substantially different from those in which the events of interest are originally observed (Gelfand et al., 2001).

An alternative technique for the analysis of disease incidence in urban areas is based on the so called ‘double kernel density’ (DKD) approach (Portnov et al., 2009; Kloog et al., 2009; Zusman et al., 2012). This technique is a nonparametric method that calculates the areal density of ‘health outcome events’ within a given search radius from a *moving* ‘target’ point and then normalizes the obtained event density by the density of background population in which the events of interest occurred (Portnov et al., 2009). The main advantage of this technique, compared to the zonal approach, is that DKD estimates do not depend on *a priory* geographic delineations of statistical zones and geographic subdivisions and help to produce a smooth and continuous disease ‘risk surface’ for the entire study area, which can subsequently be used to calculate response variables for a multivariate statistical analysis (Carlos et al., 2010; Kloog et al., 2009; Portnov et al., 2009).

In the case study, described in the following subsections, we aim to investigate the association between air pollution and cancer morbidity by comparing two analytical techniques: the traditional zonal approach (Boyle and Parkin, 1991; Curtin and Klein, 1995) and more recently developed Double Kernel Density (DKD) tools (Portnov et al., 2009; Zusman et al., 2012; Chakraborty, 2012). The study's results can be beneficial for disease risk analyses in other urban areas, in which relatively small numbers of new cancer cases are recorded annually or only few internal subdivisions are available for incidence rate calculation and comparison.

2. Background studies

2.1. Air pollution and cancer

Major metropolitan areas are often characterized by complex and confound associations between mortality and air pollution (Jarrett et al., 2013). Although in recent years, increasing public awareness about air pollution and improved air pollution