Alternative Methods of Analysis of Event Point Datasets in Epidemiological Research

Ofer Steinberg

Supervised by: Prof. Boris A. Portnov, Faculty of Management, University of Haifa

Prof. Khaled Karkabi, Clalit Healthcare Services & Faculty of Medicine, Technion

Abstract

Morbidity data in epidemiological research often represent residential locations of disease patients and have no other attributes. Statistical methods available for the analysis of such event point datasets are limited. In this study, we compare several analytical techniques, applying them to the analysis of childhood asthma in the Greater Haifa Metropolitan Area (GHMA) in Israel. Analytical techniques, which we compare, include: ordinary least square (OLS) analysis of zonal estimates, binary logistic regressions of disease events, Bayesian Model Averaging (BMA), Poisson regressions, and more recently developed Double Kernel Density (DKD) tools. Our objective is to determine which analytical approaches help to gain more information on environmental factors affecting local morbidity patterns.

The data of the residential locations obtained from 'Clalit Healthcare System' and include 24,115, 6-14 yo children (2.62% of whom were diagnosed with asthma). We mapped the children's residential addresses using ArcGIS tools and then linked each residential address to neighborhood socio-demographic attributes, air pollution averages, proximities to industrial areas, main roads, etc. Next, was analyzed the association between asthma incidence rates and the aforementioned explanatory variables, using different investigation techniques.

Our results show that the logistic regressions and zonal estimates detected no significant association between asthma incidence rates and environmental attributes ($p>0.1$), the DKD method helped to identify several environmental factors as significantly associated with the observed morbidity patterns. Poisson regressions and BMA provided less significant results.

In spite of the significance of the DKD, we explained high sensitivity of that method, compared to other analytical techniques, by the fact that it does not depend on a predefined delineation of geographic areas and helps to generate a continuous disease risk surface covering the entire study area. We recommend this method for the analysis of event point datasets in studies, in which the number of statistical areas available for comparison and aggregation is small and recorded morbidity events are relatively rare.