Space-Time Cluster Analysis: Application of Healthcare Service Data in Epidemiological Studies

Joseph M. Woodside, Cleveland State University, USA
Iftikhar U. Sikder, Cleveland State University, USA

ABSTRACT

Spatial epidemiological approach to healthcare studies provides significant insight in evaluating health intervention and decision making. This article illustrates a space-time cluster analysis using Kulldorff’s Scan Statistics (1999), local indicators of spatial autocorrelation, and local G-statistics involving routine clinical service data as part of a limited data set collected by a Northeast Ohio healthcare organization (Kaiser Foundation Health Plan of Ohio) over a period 1994—2006. The objective is to find excess space and space-time variations of lung cancer and to identify potential monitoring and healthcare management capabilities. The results were compared with earlier research (Tyczynski & Berkel, 2005); similarities were noted in patient demographics for the targeted study area. The findings also provide evidence that diagnosis data collected as a result of rendered health services can be used in detecting potential disease patterns and/or utilization patterns, with the overall objective of improving health outcomes.

Keywords: Cluster Analysis, Healthcare Service Data, Spatial Autocorrelation, Spatial Data

INTRODUCTION

The increasing demand for health data analysis in spatial and temporal scale has made emerging technologies such as Geographic Information Systems (GIS) an essential tool for healthcare information systems. In healthcare settings application of such new technology are proving useful in the analysis of health data and planning of healthcare services (Pfeiffer, Robinson, Stevenson, Stevens, Rogers, & Clements, 2008). The ability of GIS to manage and retrieve georeference data has demonstrated its value in the integration of complex epidemiological models through visualization of spatial and temporal relationships. This has been recognized by the World Health Organization (WHO):

Geographical information systems (GIS) provide ideal platforms for the convergence of disease-specific information and their analyses in relation to population settlements, surrounding social and health services and the natural environment. They are highly suitable for analyzing epidemiological data, revealing trends and interrelationships that would be difficult to discover in tabular format. Moreover GIS allows policy makers to easily visualize...
problems in relation to existing health and social services and the natural environment and so more effectively target resources. (World Health Organization, 2008)

Geographical analysis is not only important for the identification of patterns of healthcare outcomes it also offers insight into understanding the association or linkage to political processes and policy makers (Cromley, 2002; Gatrell, 2002). Health data from managed health care organizations offers the opportunity to analyze unusual geographical patterns of disease. Routine, aggregated healthcare data stored in health systems can be utilized to identify disease clusters or utilization patterns. Recently methods have been sought to further improve identification within case and disease management programs.

The real world clinical service data stored in healthcare information systems provides opportunity to analyze spatio-temporal patterns at finer granularity. The investigation of space and space-time epidemiological patterns often gives rise to the explanation of factors that might create an adverse health condition. This study uses routine, aggregated service data to find excess space and time variations in rendered services where the primary diagnosis was lung cancer. From the healthcare management point of view, if clusters are detected and explanatory factors linked, this understanding allows for better patient care, i.e. serving a particular population with targeted specialists, and preventing spread of disease amongst populations. This research aims to study different clustering methods of the spatial and spatio-temporal patterns of lung cancer particularly for routine clinical service data collected by a Northeast Ohio healthcare organization (Kaiser Foundation Health Plan of Ohio) over a period from 1994 – 2006.

**METHODOLOGY: SPACE-TIME CLUSTER ANALYSIS**

The question whether diseases such lung cancer or breast cancer are spatially clustered is an active research area (Lawson, Biggeri, Böhning, Lesaffre, Viel, & Bertollini, 1999; Laywson & Denison, 2002; Marshall, 1991; Tango & Takahashi, 2005). Since the detection of spatial and temporal patterns of clusters of lung cancer is sensitive to the clustering algorithm, it is difficult to evaluate results from a single method (Jacquez & Greiling, 2003). Currently, most of the comparative analysis of disease clusters depends on simulated data (Ozonoff, Bonetti, Forsberg, & Pagano, 2005). Tycznski developed a broad atlas of cancer in Ohio which involved a “smoothing” method where weighted averages of cancer per county were calculated versus geographic location of patients with cancer at the time of diagnosis. However, the clusters are generated by considering only spatial aspect. The temporal characteristics of the cluster are not reported (Tyczynski, et al., 2005).

In recognition of the usual epidemiological definition of cluster, this study adopts the formal definition of cluster which refers to the pattern of location of disease cases, relative to the pattern of non-cases (Wakefield, Kelsall, & Morris, 2000). In principle, since the cases are more clumped than non-cases, the difference between the two patterns is statistically recognizable. Intuitively, a cluster is an excess value which exceeds the normal value for the space and/or time. The closer a cluster population is defined, the excess value will be greater for the cluster population, and the significance will be greater. The closer a cluster population is defined, the greater the excess value will be for the cluster population, and the significance will be greater. Initial assessments of clusters include reviews of cases, boundaries of space and time, estimated number of cases, estimates of standardized mortality ratios, statistical significance, and public communication. Cluster analysis has been frequently used to identify occurrence of morbidity or unusual localized trends in disease patterns (Alexander, 1992).

A considerable amount of research in temporal and spatial context in ‘scan’ statistics has been invested in identifying disease clusters. The theory has been successfully applied in a wide variety of epidemiological studies for cluster
Related Content

An Architectural Solution for Health Information Exchange
[www.irma-international.org/article/an-architectural-solution-for-health-information-exchange/181318/](www.irma-international.org/article/an-architectural-solution-for-health-information-exchange/181318/)

Perceived Level of Benefits and Risk Core Functionalities of an EHR System
[www.irma-international.org/chapter/perceived-level-benefits-risk-core/22129/](www.irma-international.org/chapter/perceived-level-benefits-risk-core/22129/)

Factors Motivating the Acceptance of New Information and Communication Technologies in UK Healthcare: A Test of Three Models
[www.irma-international.org/article/factors-motivating-acceptance-new-information/2192/](www.irma-international.org/article/factors-motivating-acceptance-new-information/2192/)

Managing Paramedic Knowledge for Treatment of Acute Myocardial Infarction
[www.irma-international.org/chapter/managing-paramedic-knowledge-treatment-acute/13019/](www.irma-international.org/chapter/managing-paramedic-knowledge-treatment-acute/13019/)

Role of Data Mining and Knowledge Discovery in Managing Telecommunication Systems
[www.irma-international.org/chapter/role-data-mining-knowledge-discovery/47118/](www.irma-international.org/chapter/role-data-mining-knowledge-discovery/47118/)