Accident Analysis and Prevention 95 (2016) 236–249

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Land use and traffic collisions: A link-attribute analysis using Empirical Bayes method

Bruce Zi Yang*, Becky P.Y Loo

Department of Geography, The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 4 March 2016
Received in revised form 30 May 2016
Accepted 4 July 2016

Keywords:
Public health
Hazardous road locations
Land uses

ABSTRACT

Road traffic collisions represent one of the major public health problems among the leading causes of deaths globally. This paper examines several approaches in detecting hazardous road locations, and discusses the spatial distribution of these locations as well as their relationships with different land uses in Hong Kong. Two most commonly used methodologies in detecting hazardous road locations are used: the hot spot and hot zone methodologies. Both methodologies are performed using raw collision count, excess collision count and Empirical Bayes (EB) estimations. The EB estimation uses land use characteristics near the road network in defining the reference groups. Finally all the approaches are compared by a test to assess their stability. The results show that for different hazardous road location detection methodologies, the best fit estimation methods on sites are different. The results confirm some land use impacts in previous studies, and suggest some further patterns on road safety. The findings are useful in understanding the complex interrelationships between land use and road safety, and in facilitating planners and policy makers to build safer cities.

1. Introduction

Road safety is a major public health issue since traffic collisions represent one of the leading causes of death globally (World Health Organization, 2013). As an important step in improving road safety, the detection of hazardous road locations has been studied for several decades. Currently, the most commonly used methods of hazardous road location detection are the hot spot and hot zone methodologies. The hot spot methodology, which is widely used by traffic and/or police authorities around the globe, considers road segments or junctions individually. If the number of traffic collisions at a site is above the critical value, a hot spot is detected (Elvik, 2008b; Maher and Mountain, 1988; Meuleners et al., 2008). Although the thresholds vary in different countries, the ways to define hot spots are quite similar: find the most dangerous sites on the road network that has a collision measure (usually collision count or collision rate) above a critical value.

Apart from the hot spot methodology, an alternative hot zone methodology (Flahaut, 2004; Loo, 2009) has been found to be useful in detecting hazardous road locations in recent years. The hot zone methodology considers the network or spatial continuity as well as the collision measure to identify a hazardous location: multiple sites can form a hot zone if there are at least two continuous segments with high collision counts, though lower than the traditional critical value used for detecting hot sites (Loo, 2009; Loo and Yao, 2013; Yao et al., 2015). But research in the hot zone methodology is not as rich as that in the hot spot methodology (Yao et al., 2015).

There is no consensus on which of these two methodologies is better, but it is believed that they are complementary and have respective advantages for different research purposes. For instance, Loo (2009) found that the hot zone methodology is more powerful in detecting hazardous road locations on expressways and in rural/suburb areas with fewer intersections. In contrast, the hot spot methodology may be more suitable for dense urban areas with small street blocks and frequent road intersections.

Regardless of the approach used, it is understood that pure collision counts cannot represent the dangerousness of a site very well since there is randomness in the occurrence of traffic collisions in each year. Thus, many other collision measures have been proposed to represent "dangerousness". They include the collision rate, excess collision count, and Empirical Bayes (EB) estimation. Among them, the EB estimation (Hauer, 1997, 2001) was considered to be more reliable because it makes use of both the performance on the site and that on similar sites (that is, the reference population) to mediate the randomness of collision counts in a specific year (Elvik,

* Correspondence to: Room 1023, 10th Floor, The Jockey Club Tower C, Centennial Campus, Pokfulam Road, Hong Kong.
E-mail addresses: bzyang@connect.hku.hk (B.Z. Yang), bpyloo@hku.hk (B.P.Y Loo).

http://dx.doi.org/10.1016/j.aap.2016.07.002
0001-4575/

2008a; Li et al., 2008; Miranda-Moreno et al., 2007; Montella, 2010; Persaud and Lyon, 2007).

Up to date, most previous studies have only considered the engineering aspects of the road segment, such as the type of junctions, segment length or traffic volume, in calculating the expected collision count from the reference population in EB estimation. For example, Cheng and Washington (2008) used a safety performance function based on traffic volume; Persaud et al. (2010) used a function of both the road length and traffic volume. When people's activities are also considered, the land use categories could also be helpful in defining the reference population since land use is found to be among the most important environmental factors affecting road safety, especially in relation to pedestrian collisions (Dissanayake et al., 2009; Graham and Glaister, 2003; Loukaitou-Sideris et al., 2007; Priyantha Wedagama et al., 2006; Pulugurtha and Repaka, 2008; Pulugurtha and Sambhara, 2011; Wier et al., 2009).

In the 1990s, some scholars started to directly examine the relationship between land use and road safety. For example, Levine et al. (1995a,b) observed that traffic collisions are more likely to occur in employment centers than residential areas in Honolulu, Hawaii. Petch and Henson (2000) found out that the risk of children involved in a pedestrian-vehicle collision is hugely different in urban and rural areas, with the former being substantially higher. They also suggest that sub-district level analysis is essential to understand the distribution of child pedestrian-vehicle collisions. Kim and Yamashita (2002) further suggest that the largest proportion of vehicle–vehicle collisions was found near residential lands, with the second largest proportion in commercial areas. But when the rate of traffic collisions per acre of land is considered, the rate for commercial use is almost 4 times that for residential use. Although the study linked 10 years of collision data to only 1 year of land use information, the results are still enlightening since land development was quite slow in the study area. Wang and Kockelman (2013) argue that it is the mixture of residential and commercial land uses that are related to pedestrian-vehicle collisions. Elias and Shiftan (2014) found two zones in their study area to be the most dangerous for pedestrians: one is with mixed land use of residential, commercial and public buildings; the other is characterized by high residential density and low socio-economic status. In Hong Kong, Ng et al. (2002) examined 27 land use variables and their relationship with traffic risk, including area of different land use categories, number of facilities, etc.

Meanwhile, there are three major problems in previous road safety studies related to land use. The first one is that most previous studies are based on zonal analysis only. Zonal analysis, however, aggregates the collision statistics and removes the variability within the same zone. Aggregated traffic collision data at the zonal level can be integrated with land use data more easily (as both features are polygons in a GIS environment). But it does not consider the distribution of road networks and does not provide detailed location information for further hazardous road location detection. As all spatial heterogeneity within the zone is removed, a wrong impression that the zone is uniformly dangerous (or safe) may result. The larger the zone is, the greater the problem is. The second one is the representation of land use and its linkage to road safety: land use is described in a specific or highly selective manner, including the number of driveways of various types (Ivan et al., 2000), land use entropy (Wang and Kockelman, 2013), employment density, commercial, parks, number of schools, etc. (Miranda-Moreno et al., 2011) or part of an index (Cottrill and Thakuriah, 2010) instead of comprehensively based on all land use categories. The linkage between land use and collisions has similar problems: for example, only land uses near home (Elias and Shiftan, 2014) or at the nearest junction of each collision (Kim and Yamashita, 2002) are used in respective studies. We would argue that these are not good ways to link land uses to collisions because the former can only influence part of the trips, that is, home-based trips only; and the latter ignores the fact that traffic on the road network is affected not just by the closest land use to the nearest junction of a traffic collision. The third problem is that, there are some research papers on understanding the relationship between land uses and traffic collisions through modeling the former as explanatory factors for the latter, but there are very few studies on applying the knowledge about the land use impact on road safety in detecting hazardous road locations.

This paper tries to address these three research gaps. For the first one, we use the link-attribute approach based on the road network (that is, lines or one-dimensional space) instead of zonal analysis (that is, areas or two-dimensional space) is used to measure the land use attributes. For the second one, a systematic examination of all land use categories in Hong Kong and their combinations are considered. For the third one, the land use impact is incorporated into the EB method that is used in the estimation of hazardous road locations. Taking advantage of multiple years of road collision and land use data in Hong Kong, this paper will discuss the results and patterns of both hot spots and hot zones identified using different methods.

The organization of paper is straightforward. After the Introduction, a description of the methodology will then be given to illustrate our research design. Finally, the results of the analysis, spatial patterns and sensitivity tests will be presented and discussed.

2. Methodology

2.1. Study area and data

Three districts in Hong Kong are chosen as the study area: Wan Chai (on Hong Kong Island), Yau Tsim Mong (in Kowloon) and Tai Po (in the New Territories). Some basic statistics in these three districts are shown in Table 1. According to the numbers in the table, there are several differences among the three districts. First, according to the area, road network length and population, the developing intensity in the three districts are different. Yau Tsim Mong has the largest population, longest road network length and the least area. This is because the entire district is in the urban area of Hong Kong. Tai Po has the largest area (3.3% of total land area in Hong Kong), which is about three times the area of the other two districts combined, but also has the shortest road network length. This is because it is outside the urban areas in Hong Kong in the New Territories. The second difference is the possible vulnerable population in each district. Here we summarize the percentages of children (age under 14) and seniors (age over 65) in the total population in each district. The percentage of children in the total population in each district does not vary a lot from the average percentage in Hong Kong (11.6%). But the percentage of seniors varies a lot. Wan Chai has over 15% of population that are over 65 years old, but Tai Po only has 10.2%. Another noteworthy difference is that the demographics of the population in three districts are different. For example, the proportion of non-student population aged 20 and over having attained post-secondary education in Tai Po (24.1%) is close to the average in Hong Kong (25.9%). But the proportion in Yau Tsim Mong is much higher (31.8%) while the proportion in Wan Chai is even higher (46.2%). The median monthly household income in Tai Po (HKD 22,340) and Yau Tsim Mong (HKD 22,070) are slightly above the median in Hong Kong (HKD 20,500), but much lower than the figure in Wan Chai (36,150). The average domestic household sizes in three districts are roughly similar but there are more larger families living in Tai Po.