JOURNAL OF TRAFFIC AND TRANSPORTATION ENGINEERING (ENGLISH EDITION) 2016; X (X): 1–10

Available online at www.sciencedirect.com

ScienceDirect

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

Review Article

The modifiable areal unit problem in traffic safety: Basic issue, potential solutions and future research

Q3 *Pengpeng Xu* [a], *Helai Huang* [b], *Ni Dong* [c,*]

[a] *Department of Civil Engineering, The University of Hong Kong, Hong Kong, China*
[b] *Urban Transport Research Center, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China*
[c] *School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China*

ARTICLE INFO

Article history:
Received 6 August 2015
Received in revised form
24 September 2015
Accepted 29 September 2015
Available online xxx

Keywords:
Modifiable areal unit problem
Traffic safety
Regional/macro level crash
Prediction model
Traffic safety analysis zones
Review

ABSTRACT

This study fully addressed the modifiable areal unit problem (MAUP) that was well-known in geography but generally ignored by safety analysis. The basic issue of MAUP was introduced firstly with a case study to explicitly demonstrate the existence of the problem in macro level crash modeling, and then four potential strategies, i.e., using disaggregate data as possible, capturing spatial non-stationarity, designing optimal zoning systems, conducting sensitivity analysis to report the scope and magnitude of MAUP, were proposed and illustrated in an integrated way, followed by the future research directions. Results revealed that more efforts are desired to calibrate the state-of-art modeling technique at various levels of aggregation based on spatial homogeneity in traffic safety, transport characteristics, and demographical factors. The awareness of this problem in traffic safety domain is expected to the delineation of basic spatial units (e.g. the traffic safety analysis zones), as well as to provide new insights into the nature of MAUP in statistics and geography.

1. Introduction

Macro level crash prediction model to explain observed cross-sectional variations in crash counts using macro structural covariates becomes a fairly routine component in traffic safety research. Three essential incentives ensure its rapid development. First, many factors affecting crashes operate at a spatial scale (i.e., trip distribution and generation, land use pattern, and various demographical characteristics). It is desired to relate safety with zonal level factors. Then, there is a need for transportation agencies to regularly monitor region level safety and provide incentives to reduce the number of traffic casualties in a region's safety program. Therefore, a reliable assessment of safety is indispensable by estimating the aggregating crash potentials associated with the target road network on different spatial

* *Corresponding author*. Tel.: +86 18690740722.
E-mail addresses: pengpengxu@yeah.net (P. Xu), huanghelai@csu.edu.cn (H. Huang), dongni@home.swjtu.edu.cn (N. Dong).
Peer review under responsibility of Periodical Offices of Chang'an University.
http://dx.doi.org/10.1016/j.jtte.2015.09.010

scales. Furthermore, road safety is increasingly considered as a necessary component in transportation planning (FHWA, 2005). Regional safety prediction models have been suggested as means of incorporating safety considerations with long term transportation planning (Washington et al., 2006).

The last decade witnessed fast growing scope of scientific research in the context of regional or macro level safety analysis. Various spatial units have been employed in previous study, such as regions (Washington et al., 1999), counties (Aguero-Valverde and Jovanis, 2006; Amoros et al., 2003; Huang et al., 2010; Karlaftis and Tarko, 1998; Li et al., 2013; Miaou et al., 2003; Noland and Oh, 2004), districts (Haynes et al., 2007), English wards (Noland and Quddus, 2004; Quddus, 2008), zone improvement plan codes (Clark, 2003; Girasek and Taylor, 2010; Lee et al., 2014a, 2015a; Romano et al., 2006; Stamatiadis and Puccini, 2000), census tracts (Cottrill and Thakuriah, 2010; LaScala et al., 2000; Loukaitou-Sideris et al., 2007; Ukkusuri et al., 2011, 2012; Wang and Kockelman, 2013; Wier et al., 2009), block groups (Levine et al., 1995), traffic analysis zones (TAZs) (Abdel-Aty et al., 2011; Dong et al., 2014, 2015; Guevara et al., 2004; Guo et al., 2015; Hadayeghi et al., 2003, 2006; Hadayeghi et al., 2010; Lee et al., 2015b; Naderan and Shahi, 2010; Ng et al., 2002; Pulugurha et al., 2013; Siddiqui et al., 2012; Wang et al., 2013; Xu and Huang, 2015), enumeration districts (Noland and Quddus, 2005), and grid-based schemes (Kim et al., 2006; MacNab, 2004).

However, most of above mentioned zoning schemes have specific usage (i.e., TAZs are not delineated for traffic crash analysis but for long range transportation plans), and may not be appropriate for safety analysis. As a result, research based on these modifiable zoning schemes is probably subject to the modifiable areal unit problem (i.e., MAUP). Up to now, there has been little comprehensive investigation assessing how differently scale and zonal configurations influence statistical results of safety analysis. The identification and assessment of MAUP effects becomes a pressing issue as it may lead to unreliable and inaccurate safety estimation (Huang et al., 2013; Wang et al., 2012).

This study intends to fully address the MAUP that is well-known in geography but generally ignored by traffic safety analysts. The basic issue of MAUP is introduced. Firstly it's combined with a case study to explicitly demonstrate the existence of the problem in macro level crash modeling, and then four potential solutions are recommended in an integrated way, followed by the future research directions provided and suggested in the end.

2. Basic issue

MAUP refers to the situations that when the boundary of zones used in a spatial analysis changes, the statistical inference and interpretation derived from the zones are also different (Openshaw, 1984). Typically, it can be decomposed into two components: the scale and the zoning effects (Fotheringham and Wong, 1991; Openshaw and Taylor, 1979). While the scale effect describes the occurrence of statistical result variations using data aggregated at different levels, the zoning effect refers to the variability introduced by different zoning configurations at the same aggregation level.

The issue of MAUP occurs when the spatial zoning system used to collect or report geographical data is "modifiable" or "arbitrary" (Openshaw, 1984). For example, the divisions of zones based on political and administrative considerations (i.e., electoral wards and districts) have rare geographical meanings, and thus, the statistical inference based on these modifiable zones may be unreliable and questionable.

Evidence confirming the existence of MAUP could track back to 1930's (Gehlke and Biehl, 1934). Gehlke and Biehl (1934) found that the correlation coefficient between two variables tends to increase when the level of aggregation of census tracts is higher. Once the researchers realized that the analysis results on aggregated data are not reliable as expected, they started to explore the MAUP for a host of geographic problems. A survey of literature has revealed the issue in a variety of analytical contexts, including spatial interpolation (Cressie, 1996), linear regression analysis (Clark and Avery, 1976; Fotheringham and Wong, 1991), regional economic forecasting (Miller, 1998), ecological studies (Swift et al., 2008), spatial interaction modeling (Openshaw, 1977; Putman and Chung, 1989), travel mode choice modeling (Zhang and Kukadia, 2005), and TAZ delineation (Chang et al., 2002; Ding, 1998; Viegas et al., 2009).

Despite the inconsistency due to MAUP has been confirmed by geographers for decades, it has seemingly received relatively little attention in traffic safety analysis. Only three safety studies have explicitly mentioned by MACP effects (Abdel-Aty et al., 2013; Thomas, 1996; Ukkusuri et al., 2012). Thomas (1996) investigated the effect of systematic variations of road segment length on the statistical description of crash counts. Three distinct groups of roadway segments were found: (1) for small segments about 100 m, crash counts follow almost Poisson distribution; (2) for medium size segments (300–2000 m), crash counts have an intermediate empirical distribution; (3) for lager segments (more than 2000 m), crash counts are almost normally distributed. The empirical work of Thomas (1996) demonstrated that the definition of road segments may also affect the statistical results due to MAUP. However, it's a great pity that the author did not further investigate how this influence could be inserted into crash prediction modeling and hotspots identification.

In terms of the area-wide analysis, Ukkusuri et al. (2012) explored the influence of two levels of spatial aggregation (i.e., 180 zip codes and 2216 census tracts in New York Metropolitan area) on pedestrian crash frequency models. Results revealed that the zip code level models have fewer significant variables, and the census tracts based analysis provides more insightful and consistent statistical results than the one at the zip code level. In a later study, Abdel-Aty et al. (2013) investigated the effects of zonal variations (i.e., TAZs, block groups and census tracts) on the performance of Bayesian Poisson log-normal models with three different dependent variables (i.e., total crash counts, severe crash counts and pedestrian crash counts). Results indicated that the TAZs based models have more roadway/traffic related significant factors while block groups based models include more of the commute related variables. Nevertheless, above

Please cite this article in press as: Xu, P., et al., The modifiable areal unit problem in traffic safety: Basic issue, potential solutions and future research, Journal of Traffic and Transportation Engineering (English Edition) (2016), http://dx.doi.org/10.1016/j.jtte.2015.09.010