Accident Analysis and Prevention 86 (2016) 173–185

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Re-visiting crash–speed relationships: A new perspective in crash modelling

Maria-Ioanna M. Imprialou[a,*], Mohammed Quddus[a], David E. Pitfield[a], Dominique Lord[b]

[a] School of Civil and Building Engineering, Loughborough University, Loughborough LE11 3TU, United Kingdom
[b] Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 3136, United States

ARTICLE INFO

Article history:
Received 21 April 2015
Received in revised form 20 August 2015
Accepted 1 October 2015
Available online 11 November 2015

Keywords:
Traffic speed
Crash frequency
Crash severity
Pre-crash conditions
Multivariate Poisson lognormal regression
Multivariate spatial correlation

ABSTRACT

Although speed is considered to be one of the main crash contributory factors, research findings are inconsistent. Independent of the robustness of their statistical approaches, crash frequency models typically employ crash data that are aggregated using spatial criteria (e.g., crash counts by link termed as a link-based approach). In this approach, the variability in crashes between links is explained by highly aggregated average measures that may be inappropriate, especially for time-varying variables such as speed and volume. This paper re-examines crash–speed relationships by creating a new crash data aggregation approach that enables improved representation of the road conditions just before crash occurrences. Crashes are aggregated according to the similarity of their pre-crash traffic and geometric conditions, forming an alternative crash count dataset termed as a condition-based approach. Crash–speed relationships are separately developed and compared for both approaches by employing the annual crashes that occurred on the Strategic Road Network of England in 2012. The datasets are modelled by injury severity using multivariate Poisson lognormal regression, with multivariate spatial effects for the link-based model, using a full Bayesian inference approach. The results of the condition-based approach show that high speeds trigger crash frequency. The outcome of the link-based model is the opposite; suggesting that the speed–crash relationship is negative regardless of crash severity. The differences between the results imply that data aggregation is a crucial, yet so far overlooked, methodological element of crash data analyses that may have direct impact on the modelling outcomes.

1. Introduction

The primary objective of developing a traffic crash model is to elucidate the association between crashes and their potential contributory factors so as to formulate efficient and targeted crash mitigating measures. The accuracy of the modelling outcomes is therefore critical for inappropriate decisions to be avoided. Motorway crashes appear to have a decreasing trend, especially in western countries; however, the number of casualties is still anything but negligible (IRTAD, 2014; WHO, 2013). The question then arises: are the crash models we currently use accurate enough to develop appropriate preventive measures?

Each crash is the outcome of a unique sequence of events related to the involved driver(s), vehicle(s) and the road environment. The in-depth examination of individual crashes one-by-one though, is rarely possible due to the limited data availability. As a consequence, the crashes of a road network are usually analysed in a way that their volume is reduced while they remain informative (Lord and Mannering, 2010). The main crash aggregation method is based on topological and temporal criteria. In the so-called link-based (or segment-based) approach the counts of crashes that occurred on pre-defined road links during a certain time period are modelled against explanatory variables that represent the average conditions on each link (e.g. speed, traffic flow, road geometry). The explanatory power of these approaches, in terms of statistical methodology, has evolved over the years, reaching high levels of sophistication and offering better understanding of traffic crashes (e.g. Abdel-Aty and Radwan, 2000; Lord and Mannering, 2010; Ma et al., 2008; Mannering and Bhat, 2014). Despite the fact that the link-based approach is straightforward and simple, it is also by default linked with aggregation problems or else with the information loss that is aroused when multiple values are represented by a single measure (Black et al., 2009; Clark and Avery, 1976; Davis, 2004). This limitation may affect the models' explanatory potential, especially

* Corresponding author.
E-mail address: M.Imprialou@lboro.ac.uk (M.-I.M. Imprialou).

http://dx.doi.org/10.1016/j.aap.2015.10.001
0001-4575/

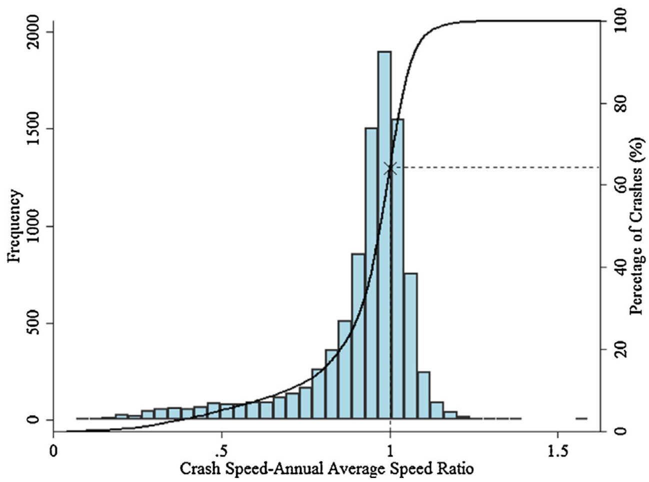

Fig. 1. Frequency and cumulative distribution of the 15-min speed at the time and the location of the crash by the annual average of the speed on this link.

for time-varying independent variables (e.g. speed, traffic volume) as their spatial and temporal variations within a link cannot be captured.

Speed is regarded as one of the main traffic related crash contributory factors (Abdel-aty et al., 2005; Elvik et al., 2004), but research findings do not confirm this unanimously. The inconsistency between the results could be partially due to the inadequacy of annual average speed to represent the speeds at which crashes actually occurred. In fact, two crashes recorded on the same link may have occurred under entirely different traffic conditions but in a link-based approach they will be both explained by the annual average speed on the link. This can be further explained by Figs. 1 and 2. Fig. 1 shows the frequency and the cumulative distribution of the ratio of the actual speed at the crash location to the annual average speed on the corresponding road link for all 2012 motorway crashes in England. Fig. 2 is the same for traffic volume. It is obvious that the ratios are considerably different from one for a high proportion of crashes (ratio = 1 means that crash speed or volume is equal with the respective annual average), confirming that the representation of time-varying measures by annual averages is rather inadequate in many cases.

This paper introduces a new crash data aggregation concept termed as condition-based approach that aims to represent in more detail the actual pre-crash conditions in order to explore the relationship between motorway crashes and their contributory factors

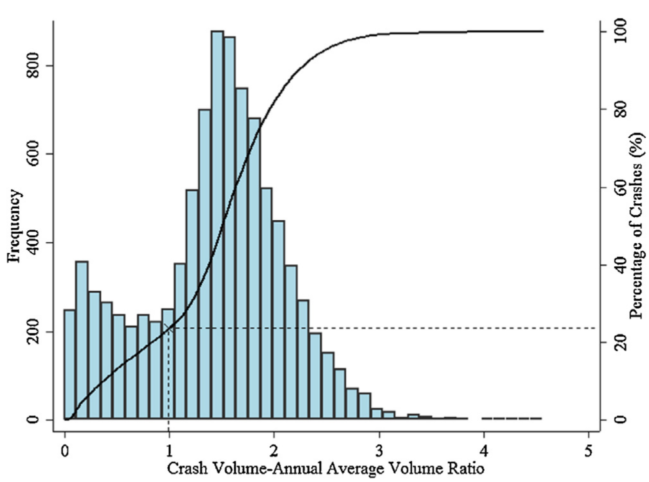

Fig. 2. Frequency and cumulative distribution of the 15-min volume at the time and the location of the crash by the annual average of the volume on this link.

such as speed, volume and geometric configuration. The grouping attribute of crashes in the proposed method is the similarity of pre-crash conditions rather than a link-level spatial relationship. In this way, crash counts are represented more precisely by explanatory variables that approximate the actual conditions enabling, possibly, improved relationships. The condition-based dataset can be modelled using multivariate Poisson lognormal regression. In order to compare the two methods with respect to their outcomes, the same data are also used to build a link-based spatial multivariate Poisson lognormal regression model.

2. Literature review

A considerable amount of literature has been published on the impact of various traffic and geometric road characteristics on link-based crash frequency. Among others speed, traffic volume, number of lanes, gradient and horizontal curvature are widely studied. From a qualitative point of view, findings show that although crash severity is positively correlated with driving speed (Clarke et al., 2010; Joksch, 1975; Kloeden et al., 1997; Pei et al., 2012), the relationship between speed and crash frequency is not equally straightforward (Aarts and Van Schagen, 2006). The early study of Solomon (1964) was the first to suggest that speed and crash frequency are not proportional but their relationship can be described by a "U-shaped" curve; an idea that was supported by several other researchers (e.g., Munden, 1967; Cirillo, 1968). Solomon's curve implies that only extremely low and high speed conditions trigger crashes. However, most of the subsequent studies find driving speeds to be linearly or exponentially related to crashes (Baruya and Finch, 1994; Fildes et al., 1991; Kloeden et al., 2002, 1997; Taylor et al., 2000). A few studies contradicted this view proposing that the speed–crash relationship is negative (Baruya, 1998; Stuster, 2004) and others reported that this relationship is statistically insignificant (Garber and Gadiraju, 1989; Lave, 1985). Some of the most recent papers that employed advanced statistical models did not find statistically significant relationships between speed and crashes (Kockelman and Ma, 2007; Quddus, 2013). Pei et al. (2012) attempted to explain the results' inconsistencies suggesting that the crash–speed relationship that is estimated by models strongly depends on the selected measure of exposure; the relationship was shown to be negative for distance-based exposure (i.e., vehicle miles travelled) but positive for time-based exposure (i.e. vehicle hours travelled).

The relationship of speed with crashes cannot be defined without taking into account the simultaneous effect of other traffic characteristics such as traffic flow (Aarts and Van Schagen, 2006) and vehicle occupancy (Garber and Subramanyan, 2001; Lord et al., 2005a). High traffic flow (represented by AADT, hourly volume, etc.) is generally considered to increase the risk of crashes (Abdel-Aty and Radwan, 2000; Chang, 2005; Milton and Mannering, 1998). On the contrary, lower flows have been also correlated with higher speed variance that is also considered to be a significant crash precursor (e.g., Garber and Ehrhart, 2000; Elvik et al., 2004). The mechanism of its impact though is not explicitly explained because of the lack of individual vehicle-level second-by-second data that would lead to reliable estimations. Instead, current studies employed relatively highly aggregated data that lead to inconclusive results (Garber and Ehrhart, 2000; Kockelman and Ma, 2007; Quddus, 2013; Solomon, 1964). Although seldom researched, vehicle occupancy ratio was found to have a non-linear relationship with the number of crashes (Garber and Subramanyan, 2001) and was also dependent on the number of vehicles involved in the crash (i.e., single- versus multi-vehicle crashes) (Lord et al., 2005a).

Road geometric design is also believed to be related with crash frequency on the roadway (AASHTO, 2010). High crash frequency is associated with high vertical grades (Anastasopoulos