Predicting the safety impact of a speed limit increase using condition-based multivariate Poisson lognormal regression

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: IMPRIALOU, M-I., QUDDUS, M.A. and PITFIELD, D.E., 2016. Predicting the safety impact of a speed limit increase using condition-based multivariate Poisson lognormal regression. Transportation Planning and Technology, 39 (1), pp. 3-23.

Additional Information:

- This is an Accepted Manuscript of an article published by Taylor & Francis in Transportation Planning and Technology on 25th November 2015, available online: http://www.tandfonline.com/10.1080/03081060.2015.1108080.

Metadata Record: https://dspace.lboro.ac.uk/2134/19642

Version: Accepted for publication

Publisher: © Taylor & Francis

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
Predicting the safety impact of a speed limit increase using condition-based multivariate Poisson lognormal regression

Maria-Ioanna M. Imprialou*, Mohammed Quddus, David E. Pitfield

School of Civil and Building Engineering, Loughborough University, Loughborough, United Kingdom

School of Civil and Building Engineering, Loughborough University, Loughborough LE11 3TU, United Kingdom, Tel: +44(0)1509 228545, email: M.Imprialou@lboro.ac.uk

*corresponding author
Predicting the safety impact of a speed limit increase using condition-based Multivariate Poisson lognormal regression

Speed limit changes are considered to lead to proportional changes in the number and severity of crashes. To predict the impact of a speed limit alteration, it is necessary to define a relationship between crashes and speed on a road network. This paper examines the relationship of crashes with speed, as well as with other traffic and geometric variables, on the UK motorways in order to estimate the impact of a potential speed limit increase from 70 mph to 80 mph on traffic safety. Full Bayesian multivariate Poisson lognormal regression models are applied to a dataset aggregated using the condition-based approach for crashes by vehicle (i.e. single-vehicle and multiple-vehicle) and severity (i.e. fatal or serious and slight). The results show that single-vehicle crashes of all severities and fatal or serious injury crashes involving multiple vehicles increase at higher speed conditions and particularly when these are combined with lower volumes. Slight injury multiple-vehicle crashes are found not to be related with high speeds, but instead with congested traffic. Using the speed elasticity values derived from the models the predicted annual increase in crashes after a speed limit increase on the UK motorway is found to be 6.2-12.1 % for fatal or serious injury crashes and 1.3-2.7% for slight injury, or else up to 167 more crashes.

Keywords: speed limit increase; speed and crashes; single vehicle crashes; multiple vehicle crashes; condition-based modelling

1 Introduction

The aim of speed limits is to maintain the equilibrium between road safety, traffic flow and energy consumption in road networks (TRB 1998; Department for Transport 2006). Although speed limits should not be considered as the target speed, a great proportion of drivers in Great Britain systematically exceed the speed limits; in 2013 47% of cars violated the 70 mph speed limit, leading to an annual 85th percentile of speed as high as 77mph (Department for Transport 2014). One could argue that these figures indicate that the current speed limit needs to be updated, as it was formerly proposed by the Department for Transport in 2011 (Department for Transport 2011a). On the other hand,
a speed limit increase raises concerns about its potential safety consequences. To assess whether a speed limit increase is sensible and appropriate it is necessary to estimate its future impact on road safety.

The relationship of speed with crashes is the key to the quantification of the impact of speed limit changes. The majority of the before-after studies report that speed limit alterations lead to proportional changes in crash frequency (Elvik, Christensen, and Amundsen 2004; Aarts and Van Schagen 2006). This effect is always attributed to the increase of average speeds on the roadway; however, when the relationship of speed with crashes is examined individually the outcomes remain rather inconclusive. The literature includes studies that have found a proportional (e.g. Kloeden, McLean, and Glonek 2002), inversely proportional (e.g. Baruya 1998a) and insignificant (e.g. Kockelman and Ma 2007) relationships between speed and crashes. The inconsistency in the results may lie in the inability of the link-based crash aggregation approaches to represent the actual conditions on the network and also the tendency to examine highly aggregated crash counts as a basis for the analysis. Link-based analyses use variables that are by default highly aggregated as they represent an entire link with one characteristic value (e.g. time-varying measures are usually represented by annual averages) (Clark and Avery 1976). In this way it is likely that the spatial and temporal variations within the link are not captured, making the representation of the pre-crash conditions practically impossible. Moreover, analysing all crash types combined may reduce the capability of models to reveal the actual crash contributory factors as those are found to vary between different crash generation processes (Geedipally and Lord 2010; D. G. Kim et al. 2007).

The aim of this paper is to explore the relationship between speed and vehicle crashes on the UK motorways in order to quantify the safety impact of a potential speed increase.
limit increase from 70 mph to 80 mph on the UK motorways. The study is based on the
development of advanced statistical models that address some of the limitations of the
link-based models and therefore lead to more accurate predictions on the changes on
crash frequency that can be expected after a speed limit increase.

This paper is organised as follows: firstly, the key empirical findings and the
current methodological approaches of the existing literature are reviewed. Next, is the
presentation of the data used for the analysis as well as their pre-processing method.
This is followed by the description of the statistical method and the results on the effect
of a speed increase on crashes by type. The final section summarises the main outcomes
of the study and some recommendations for future research.

2 Literature Review

Speed limit changes lead to proportional but comparatively moderate changes in the
average speed of road networks (Rock 1995). Literature suggests that the average speed
change equals approximately to one quarter to half of the speed limit difference (e.g.
Aljanahi, Rhodes, and Metcalfe 1999; Baruya 1998b; Elvik, Christensen, and
Amundsen 2004; Finch et al. 1994; Freedman and Williams 1992). Assuming that a
relationship between speed and crash frequency exists, a speed limit increase, if all
other factors remain unchanged, should lead to increased number of crashes on a road
network. This is confirmed by most (70.5%, weighted percentage) of the before-after
studies that examined the impact of speed raises and were included in a meta-analysis
by Elvik et al. (2004).

There have been several attempts in the literature to define a general rule for the
impact of speed limit changes; most of them are based on the combination of the
outcomes of previous studies (i.e. meta-analyses). Nilsson (2004) suggested that the
impact of a speed limit change can be quantified using a Power model that was validated in two extensive meta-analyses by Elvik et al. (2004) and Elvik (2009) where they developed several power functions suitable for impact estimation. The use of power functions is a straightforward and transferrable method however its main drawback is that the exponents provided are independent of the baseline speed that might not lead to accurate estimations. This was initially suggested by Hauer and Bonnenson (Hauer and Bonneson 2006) who addressed some of the Power Model’s limitations developing an exponential model using data from Elvik’s et al. meta-analysis ((Elvik, Christensen, and Amundsen 2004).

Meta-analyses’ results are useful for identifying general data patterns, but may be not accurate enough for predicting the effect of a speed limit increase in a particular road network as they cannot take into account area-specific characteristics (geographic, cultural etc.) that may differentiate the outcomes. Consequently, to predict the impact of a future speed limit increase on a specific road network it is necessary define the current crash-speed relationship on it. Based on the amount of the kinetic energy that is released during a collision, crashes that occur under high speed conditions are definitely more likely to lead to more serious outcomes (Joksch 1993; Aarts and Van Schagen 2006; Pei, Wong, and Sze 2012). Speed is also considered to be related with higher crash frequency. That is mainly because high speeds are related with longer stopping distances and increased probability of loss of control or other errors (Kloeden et al. 1997; Elvik, Christensen, and Amundsen 2004). Although there is a considerable amount of research on this topic, there are several points of disagreement between studies. A number of researchers suggest that speed and crash frequency are proportional (Taylor, Lynam, and Baruya 2000; Fildes, Rumbold, and Leening 1991), however others did not find any statistically significant relationship between them.
(Kockelman and Ma 2007; Quddus 2013). Additionally, a few studies contradicted the common belief, proposing that speed is inversely proportional with crashes (Baruya 1998b; Stuster 2004).

A source of variability in the research outcomes might be the information losses caused by the conventional crash data aggregation approaches that employ spatial criteria, such as the link-based method. In link-based models all the crashes that occurred on a road link during the study period are grouped and analysed together under the strong assumption that they are all related with the average traffic conditions on the link. This method is related with data aggregation problems (Black, Hashimzade, and Myles 2009; Davis 2004, Imprialou et al. forthcoming) and limited potential to represent the actual traffic conditions that are related with crashes which are very likely to be extreme (Hossain and Muromachi 2013; Pande and Abdel-Aty 2005).

Most of the studies that examine crash frequency as a function of speed employ the total number of crashes that occurred on a network. However, different crash mechanisms could be by definition related with different traffic circumstances (D. G. Kim, Washington, and Oh 2006). As a consequence, the examination of crash contributory factors to an aggregate level might distort the results of the analyses.

Researchers who studied the effects of crash contributory factors for different crash types (defined by the number of involved vehicles and/or the point of the first impact) confirm that there are indeed significant variations in the estimated coefficients between crash types (Qin, Ivan, and Ravishanker 2004; D. G. Kim, Washington, and Oh 2006; Ye et al. 2009; Bham, Javvadi, and Manepalli 2012; Geedipally, Patil, and Lord 2010; Ivan, Pasupathy, and Ossenbruggen 1999; Ivan, Wang, and Bernardo 2000). Single vehicle (henceforth: SV) crashes are found to be related with low density traffic conditions in contrast to multiple vehicle (henceforth: MV) crashes that are associated
with peak periods, higher volume and density (Ivan, Pasupathy, and Ossenbruggen 1999; Ivan 2004; Ivan, Wang, and Bernardo 2000). Therefore, the separate examination of speed with SV and MV crashes is logical and justified. A limitation of the previous studies was the use of separate regression models for each crash type. This approach ignores possible correlations between different crash types that can potentially lead to imprecise estimations (Park and Lord 2007). This can be addressed using multivariate Poisson or Poisson lognormal models that have been proposed for modelling simultaneously multiple crash categories while controlling for their correlations (e.g. Ma and Kockelman 2006; Ma, Kockelman, and Damien 2008; Aguero-Valverde and Jovanis 2009; Park and Lord 2007; El-Basyouny and Sayed 2009; Barua, El-Basyouny, and Islam 2014; Aguero-Valverde 2013). At most of the cases, multivariate models have been shown to have improved fit to the data compared to univariate models and that is why they are considered to provide more accurate outcomes (Barua, El-Basyouny, and Islam 2014; Ma, Kockelman, and Damien 2008).

This paper defines the relationship of speed with crashes on the UK motorway by developing a modelling approach that eliminates aggregation problems, which is the main limitation of the link-based models, and takes into account the correlations between the two examined crash types using multivariate Poisson lognormal regression. Crashes are grouped according to the resemblance of the prevailing traffic conditions just before their occurrence which are identified based on their geo-coded crash locations. The crash counts are divided by crash type such as single vehicle crashes and multiple vehicles crashes and by severity and are modelled simultaneously using multivariate Poisson lognormal regression in a full Bayes framework. Through the development of the crash-speed relationships, the impact of a potential average speed increase caused by a speed limit raise from 70 mph to 80 mph is quantified.
3 Data preparation

3.1 Data

The data that were synthesized for the statistical analysis were obtained from: a) the National Road Crash Database of the United Kingdom (STATS 19) (Department for Transport 2011b), b) the UK Highways Agency Journey Time Database (JTDB) (Highways Agency 2011) and c) the UK Highways Agency Traffic speed condition survey (TRACS) (Highways Agency 2008). Each of these datasets is briefly discussed below.

3.1.1 Crash Data

The examined crash data consisted of 5,606 crashes that occurred on the motorway network of England (total length approximately 3,519 km, typical speed limit 70 mph) during 2012. STATS 19 crash reports include all injury crashes and are divided into crashes with fatal, serious and slight injuries. The variables that were extracted from STATS19 for the purpose of this analysis are crash date, time, location, number of vehicles involved and vehicles’ intended direction prior to the crash. Considering the different intrinsic characteristics of collisions, crashes will be examined separately by type. Crashes were divided according to the number of involved vehicles into: a) single vehicle crashes (SV) and b) main carriageway multiple vehicle crashes (MV).

Intersection MV crashes (defined as crashes where the colliding vehicles had different intended directions) were eliminated from the analysis for two reasons. Firstly, the small number of observations (4.6% of all motorway crashes) did not permit the formation of an individual category that was suitable for count regression models. Secondly, intersection crashes could not be merged with the main carriageway multiple vehicle collisions as these types are assumed to have significantly different generation
processes. The final number of valid crash observations was 4,505 crashes\(^1\) due to

missing and/or illogical values either in the crash, traffic or geometry datasets.

Crash location is a key component for the identification of the traffic and

generic conditions just before a crash; however, the reported crash locations were less

accurate than desired. To overcome this limitation, crashes were allocated to refined and

more accurate locations based on the output of a fuzzy-logic based crash mapping

algorithm which employs some of the most common crash location information

(Imprialou, Quddus, and Pitfield 2014). The algorithm was developed for the study area

and provides 98.9\% (±1.1\%) accurate crash locations.

3.1.2 Traffic data

Traffic conditions were obtained from the JTDB that stores link-level traffic data

(obtained from inductive loops) of the UK motorway network in 15-minute intervals.

The traffic variables used here were the average speed (km/h) and the volume (vehicles)

per 15 minutes.

3.1.3 Geometric data

Road geometry was obtained by the TRACS surveys that measure road geometrical

characteristics using survey vehicles. The data used here were the radius and gradient

measured in a 10-metre span for the entire UK motorway network.

\(^1\) Out of the 4,505 crashes included in the analysis 1,060 were single vehicle (184 fatal or

serious and 876 slight) and 3,445 were multiple vehicle (302 fatal or serious and 3143

slight).
3.2 **Condition-based datasets**

In link-based approaches crashes are gathered into groups based on the location of their occurrence. As it has been stated before, this default feature of this method might influence the outcomes and restrict the explanatory potential of the models. In this paper, to overcome this limitation crashes are aggregated in an alternative way. Instead of their locations, the examined crashes are grouped according to the similarities of the traffic and geometric road conditions just before their occurrence forming a different crash count dataset termed as condition-based (see also Imprialou et al. forthcoming). To generate a condition-based dataset the combination of the crash, traffic and geometry data in a form suitable for the statistical analysis is required. Each observation of the condition-based dataset represents an individual scenario of traffic and geometric conditions. The final dataset consists of every possible condition scenario that could occur on the road network during the study period. The condition scenarios were formed by combining the following variables:

- **Speed**: Speed was divided into 50 groups of equal frequency with a 2-percentile step (e.g. $2^{nd}-4^{th}$ percentile of speed observations) and was represented as a continuous variable by the median of each group.

- **Volume**: Separately for each of the 50 speed groups, the volume was divided into 4 intervals of equal frequency (i.e. the quartiles of volume observations) resulting in a total of 200 unique values and was represented as a continuous variable by the median of each interval.

- **Curvature**: Sections with curved or straight sections (dummy variable);

- **Gradient**: Uphill, downhill or level sections (categorical variable);

- **Number of lanes**: Sections with up to two lanes per direction or over two lanes per direction (dummy variable).
Combining the above independent variables the final dataset included overall 2,400 scenarios which traffic conditions appeared with equal frequency on the study network. Following, each scenario of the dataset was matched with a number of crashes (if any) that occurred under its corresponding traffic and geometric conditions (dependent variable). To do this, each crash was individually classified to the condition scenario that described best the road circumstances prior its occurrence.

Traffic conditions on the road section where the crashes occurred were identified based on the reported date and time of the crash. The road segment that was assumed to be the most influential for the crash occurrence was considered to be equal with the length of the average stopping distance upstream of the crash location increased by 20 metres downstream to correct for minor errors related with the crash location identification (final considered segment length: 117 metres). Based on the TRACS measurements, each segment was characterised as curved or straight and uphill, downhill or level.

After the classification of crashes to specific condition scenarios, each scenario included crash counts split by crash type (i.e. SV, MV) and by severity. To control for the unequal likelihood of crash occurrences between scenarios the measure of exposure that was considered as more suitable was the average vehicle hours travelled per mile for each scenario. The descriptive statistics of the dataset can be found at Table 1.

4 Methodology
Crash counts by collision type cannot be assumed to be independent because they are subsets of the total crashes that occurred on a road network. Therefore, modelling them separately might lead to inaccurate estimations of standard errors (Park and Lord 2007). Multivariate Poisson Log Normal (MVPLN) regression is proposed for modelling categorised crash counts (e.g. by collision type or severity level) while controlling for
over-dispersion and the correlations between the categories (Park and Lord 2007; Ma, Kockelman, and Damien 2008; El-Basyouny and Sayed 2009; Aguero-Valverde and Jovanis 2009).

In a condition-based dataset with \( n \) pre-crash scenarios the number of crashes per category can be considered to follow a Poisson distribution with a lognormally distributed parameter:

\[
y_{ik} \sim \text{Poisson}(\lambda_{ik}), \quad i=1,2,\ldots,n, \quad k=1,2,\ldots,K
\]  

Where \( i \): index of observation, \( k \): index of crash category, \( y_{ik} \): observed number of crashes of the category \( k \) for the \( i^{th} \) observation and \( \lambda_{ik} \): the expected mean of crashes of category \( k \) for the \( i^{th} \) observation. The link function of the model is:

\[
\ln(\lambda_{ik}) = \beta_{k0} + \sum_{m=1}^{m} \beta_{km} X_{ikm} + \ln(e_i) + \epsilon_{ik}
\]

Where \( \beta_{k0} \): intercept for category \( k \), \( \beta_{km} \): coefficient of the \( m^{th} \) explanatory variable and category \( k \), \( X_{ikm} \): value of the \( m^{th} \) explanatory variable for the \( i^{th} \) observation and category \( k \), \( e_{i} \): exposure variable and \( \epsilon_{ik} \): unobserved heterogeneity for the \( i^{th} \) observation and category \( k \). \( \epsilon_{i} \) is multivariate normally distributed so as to control for the correlations within the unobserved heterogeneity:

\[
\epsilon_{i} \sim \text{MVN}(\mathbf{0}, \Sigma), \quad \Sigma = \begin{pmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1K} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{1K} & \sigma_{K2} & \cdots & \sigma_{KK}
\end{pmatrix}
\]
Where $\Sigma$ is the variance-covariance matrix (i.e. precision matrix) of the unobserved heterogeneity.

The models’ parameters were estimated using Markov chain Monte Carlo (MCMC) in a Bayesian framework. The prior distribution for $\beta$ was multivariate normal and the conjugate prior distribution of the precision matrix was Wishart as it has been suggested in the literature (e.g. Aguero-Valverde and Jovanis 2009; Ma, Kockelman, and Damien 2008; Park and Lord 2007):

$$\beta \sim \text{MVN}(\beta_0, R_{\beta_0})$$

$$\Sigma^{-1} \sim \text{Wishart}(R, d)$$

Where $\beta_0$, $R_{\beta_0}$, $R$ are non-informative hyperparameters and $d$ represents the degrees of freedom (i.e. $d=k$).

The model presented in equation (8) was applied separately to: (i) all crashes split by crash type (henceforth: CT) (i.e. all SV and all MV crashes), (ii) SV crashes disaggregated by severity (henceforth: SV_sev) (i.e. SV crashes with killed and serious injuries combined and slight injury crashes), (iii) MV crashes split by severity (henceforth: MV_sev) (i.e. MV crashes with killed and serious injuries combined and slight injury crashes). The actual functional form of the relationships of speed and volume with crashes is unknown (Qin, Ivan, and Ravishanker 2004). In order to examine the validity of the assumption that the relationship of crashes with speed and

---

2 Fatal and serious crashes were combined into one category due to the very low number of observations for fatal crashes (i.e. 25 out of 1060 for SV and 36 out of 3445).
volume is necessarily linear, 18 different specifications\textsuperscript{3} for speed and volume were
tested. The models that are presented in the following section have the best fitting
specification among all the examined specifications (based on the Deviance Information
Criterion).

5 Results and discussion

5.1 Modelling outcomes

Three models were estimated with the MCMC method using WinBUGS software
(Spiegelhalter et al. 2003) that is suitable for multivariate models with the full Bayesian
approach. The estimations were derived from 50,000 iterations of two chains with a
burn-in sample of 20,000. Convergence was visually detected by observing the trace
plots of the estimates. The best fitting specification was different for each of the three
models; more specifically CT was best described by a linear speed, the natural
logarithm of volume and an interaction term (i.e. Speed*Volume) (i.e. specification 12),
SV\textsubscript{sev} by the square of speed and the logarithm of volume (i.e. specification 3) and
MV\textsubscript{sev} by linear speed, squared volume and an interaction term (i.e. specification 11).
Tables 2-4 present the coefficient estimates for CT, SV\textsubscript{sev} and MV\textsubscript{sev} respectively.
In order to clarify the outcomes, especially for the models that have interaction terms,

\textsuperscript{3} 1. speed, volume; 2. speed, volume\textsuperscript{2}; 3. speed, ln(volume); 4. speed\textsuperscript{2}, volume; 5. speed\textsuperscript{2},
volume\textsuperscript{2}; 6. speed\textsuperscript{2}, ln(volume); 7. ln(speed), volume; 8. ln(speed), volume\textsuperscript{2}; 9. ln(speed),
ln(volume); 10. speed, volume, speed*volume; 11. speed, volume\textsuperscript{2}, speed*volume; 12.
speed, ln(volume), speed*volume; 13. speed\textsuperscript{2}, volume, speed*volume; 14. speed\textsuperscript{2}, volume\textsuperscript{2},
speed*volume; 15. speed\textsuperscript{2}, ln(volume), speed*volume; 16. ln(speed),volume,
speed*volume; 17. ln(speed), volume\textsuperscript{2}, speed*volume; 18. ln(speed), ln(volume),
speed*volume;
Figures 1a, 1c, 4a, 4c, 7a and 7c depict the relationship of speed with crash rates for 3
distinct volumes. Similarly, Figures 1b, 1d, 4b, 4d, 7b and 7d represent the relationship
of volume with crash rates for three different speeds. Figures 2, 3, 5, 6, 8 and 9 show the
relationship of crash rates with speed and volume combined in a 3D format.

Comparing the coefficient values of the SV with those of the MV (Table 2,
Figure 1, Figure 2 and Figure 3) it is clear that the two crash types tend to occur under
significantly different traffic conditions on the roadway. SV collisions increase while
speed increases and volume decreases; in other words, they tend to occur more
frequently at lower density conditions. On the other hand, MV collisions seem to be
related with lower speeds and higher volumes and consequently with more intense
traffic. In principle, this outcome, that is in line with the existing literature (Ivan, Wang,
and Bernardo 2000; Ivan, Pasupathy, and Ossenbruggen 1999; Ivan 2004), re-confirms
that modelling crashes by type is advantageous as it can be more informative about the
circumstances that particular crashes types occur. However, interpreting these results
without looking at the effect of speed on different severity levels by crash type might be
misleading, especially for the case of MV crashes.

The frequency of SV crashes is independent of the severity of their outcomes
and increases proportionally with speed (Table 3, Figure 4, Figure 5 and Figure 6). The
coefficients of speed for KSI and SL crashes show that, as expected, higher speed is
also related with more serious injuries. Traffic volume has exactly the opposite effect to
SV crashes, as the highest crash rate is observed at lower volumes, a finding that is
consistent with previous studies (e.g. Qin, Ivan, and Ravishanker 2004; Ivan 2004). The
results of the SV_sev model are explainable as SV crashes are probably the most speed-
related crash type. SV crashes are associated with loss of control, alcohol or drug
impaired drivers, risk-taking actions, fatigue and sleepiness (Xie, Zhao, and Huynh
They usually occur during off-peak times and especially at night time when density is at low levels and vehicle encounters are less likely (Ivan, Pasupathy, and Ossenbruggen 1999). Thus, SV crashes are expected to be more and more serious after an average speed limit increase.

Despite the initial findings of the CT model, MV crashes are found to have different relationships with traffic conditions when they are disaggregated by severity level (Table 4, Figure 7, Figure 8 and Figure 9). The outcome of the CT model reflects mainly the relationship of the majority of MV crashes that accounted for slight injuries. Slight injury MV crashes seem to have a negative relationship with speed while their relationship with volume can be described by a U-shaped curve (Figure 9) meaning that this crash type tends to occur mainly at very low and very high volume conditions. This outcome can be explained considering the characteristics of the two main collision types of MV same direction crashes: side and rear-end collisions. Side impacts are more likely to occur during overtaking manoeuvres that are more frequent under lower density conditions. On the contrary, rear-end collisions are linked with more dense conditions. This is consistent with previous findings that suggest that high traffic intensity and peak hours are related with more MV crashes (Ivan 2004; Ivan, Pasupathy, and Ossenbruggen 1999). Fatal and serious injury MV crash rate, though, is generally proportional to speed apart from very high volume conditions. This is in line with existing studies that suggest that crashes with serious impact tend to be positively related with speed (e.g. Aarts and Van Schagen 2006). This result may reflect the

4 When the 15-minute volume per lane is above 278 vehicles (estimated based on the slope equation for speed). This corresponds to volumes higher than the 83rd percentile of all observations.
differences of the two major same direction crash types but it might also be the effect of
the merger of fatal and serious crashes into one category. If it would be possible to
estimate a separate model for fatal crashes, speed would have probably shown a
positive relationship with fatal crashes.

Road geometry was mainly found to have similar impact on both SV and MV
crashes, as the coefficients were consistent for all types and severities. Locations with
curvature and steep horizontal alignment (especially downgrades) tend to concentrate
more crashes, a finding that is similar with previous literature (Milton and Mannering
1998; Abdel-Aty and Radwan 2000; Anastasopoulos and Mannering 2009). Roads with
more lanes are found to be more related with crash occurrences too. This can be
explained especially for MV crashes by the fact that wider roads are more prone to lane
changing that is related with increased and potentially dangerous vehicle encounters
(e.g. Chang 2005).

5.2 Impact estimation

Apart from explaining the relationship of crashes with traffic and geometry
related variables, the developed models are employed to predict the impact of speed
limit changes on traffic crashes. Using the elasticity of crashes with respect to speed it is
possible to estimate the expected changes in the number of crashes by type and severity
as a result of a speed limit increase. According to existing literature the average speed
on a road is expected to rise by 25% to 50% of the amount of the speed limit increase
(e.g. Rock 1995; Finch et al. 1994). This means that if the speed limit of UK motorways
increases from 70 mph to 80 mph (i.e. 10 mph) the average speed would be expected to
increase by 2.5 mph to 5 mph. However, it is not clear how this change would affect the
speed distribution of the network. A speed limit change could cause a uniform shift to
the speed distribution, or it could cause a more significant increase at higher speed
conditions than at the lower ones. Considering that low speeds are normally related with traffic congestion, the second case is likely to be more representative. Since it is not possible to predict the form of the new speed distribution, the elasticity values that are presented in this paper are estimated based on the expected changes on the average speed. The equation of the mean elasticity of the $m^{th}$ variable of the $k^{th}$ category is:

$$Elasticity = \frac{\partial E(y|x_{mk})}{\partial x_{mk}} \frac{x_{mk}}{y}$$  \hspace{1cm} (12)

Table 5 shows the mean elasticity of speed and the estimated minimum and maximum percentage of increase for SV and MV crashes based on the outcomes of the SV_sev and the MV_sev models respectively. As discussed, a 10 mph increase in speed limit would result in a 3.86% in average speed rise (i.e. the average speed 64.7 mph would at least increase by 2.5 mph). Given that the mean elasticity of crashes with respect to speed is 2.595 for SV KSI crashes (see Table 5), the corresponding increase in these crash type would be at least 10.02% (i.e. 3.86*2.595) as shown in Table 5. In a similar manner, SV SL and MV KSI crashes would have an increase of 6.14% (i.e. 3.86*1.590) and 3.42% (i.e. 3.86*0.886) respectively. The speed elasticity for slight injury MV crashes was chosen not to be presented here. As the relationship of speed with this crash type is negative, the elasticity of speed is a negative, too. Having no evidence to support that a speed limit increase can be associated with decrease in particular types of crashes and to keep the results conservative it is considered that the number of MV crashes that lead to slight injuries will not change.

Assuming that all other variables remain the same, single vehicle KSI are expected to increase by 10.0%-20.1% after the first year of implementation of the measure and for SL this number will fluctuate from 6.2% to 12.39%. This means that after a speed limit increase there will be 73-146 more SV occurrences on the UK
motorway. The increase of MV crashes will be from 3.4%-6.9%\% equivalent to 11-21 more MV crashes. The overall predicted increase due to the anticipated average speed raise for all KSI and SL crashes will reach 6.2%-12.1 \% and 1.3\% -2.7\% respectively, indicating that a change on the current speed limit will clearly have a considerable and adverse impact on road safety.

6 Conclusion

Changes in speed regulation laws lead to changes of the traffic conditions that might affect the levels of safety on road networks. To predict the impact related with such measures it is necessary to understand the relationship of speed with crashes on the examined network. This paper explores the relationship of speed with single vehicle and same direction multiple vehicle crashes on the UK motorways so as to evaluate the effect of a potential 10 mph increase of the current 70 mph speed limit. The speed-crash relationship is described through three condition-based multivariate Poisson lognormal regression models that provide different coefficients by type and severity of the crashes respectively.

The results of the models show that speed is positively related with all single vehicle crashes and the fatal or serious multiple vehicle crashes but negatively related with multiple vehicle crashes with slight injuries. This outcome suggests that the UK motorway is likely to have 6.2-12.1 \% more fatal or serious crashes and 1.3-2.7\% more slight injury crashes during the first year of the new speed limit implementation, confirming the concerns about the appropriateness of this measure. Taking into consideration that speed limit increases tend to be linked with increases in average speeds on contiguous roads (i.e. spillover effect), the overall crash rise might be even higher than the estimated. As a consequence, a speed limit increase, in the absence of
new and effective preventive measures, does not seem to be a reasonable idea assuming that road safety is one the first priorities for policy makers and other stakeholders. Instead, a reduction of the current number of speed limit violations, that could be achieved through improvements in enforcement, would be beneficial as it would lead to a decrease of crashes and the severity of their outcomes.

7 References


Table 1: Descriptive statistics of the condition-based dataset

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>All_SV</td>
<td>0.442</td>
<td>0.833</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>KSI_SV</td>
<td>0.077</td>
<td>0.292</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>SL_SV</td>
<td>0.365</td>
<td>0.730</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>All_MV</td>
<td>1.435</td>
<td>3.719</td>
<td>0</td>
<td>80</td>
</tr>
<tr>
<td>KSI_MV</td>
<td>0.203</td>
<td>0.548</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>SL_MV</td>
<td>1.310</td>
<td>3.461</td>
<td>0</td>
<td>75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed (mph)</td>
<td>64.692</td>
<td>8.301</td>
<td>31.771</td>
<td>82.282</td>
</tr>
<tr>
<td>Volume per lane (measurement interval 15 minutes)</td>
<td>147.92</td>
<td>113.694</td>
<td>11.000</td>
<td>443.000</td>
</tr>
<tr>
<td>Curve</td>
<td>0.500</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Straight (reference)</td>
<td>0.500</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Uphill</td>
<td>0.333</td>
<td>0.472</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Downhill</td>
<td>0.333</td>
<td>0.472</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Level (reference)</td>
<td>0.333</td>
<td>0.472</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Lanes above 2</td>
<td>0.500</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Lanes up to 2 (reference)</td>
<td>0.500</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Vehicle hours per mile (Exposure)</td>
<td>7.039</td>
<td>6.735</td>
<td>0.526</td>
<td>45.934</td>
</tr>
</tbody>
</table>
Table 2: Coefficient estimates of the multivariate Crash Type (CT) model for all crashes (i.e. Single-Vehicle and Multiple-Vehicle)

<table>
<thead>
<tr>
<th></th>
<th>Single Vehicle</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>MC error</td>
<td>2.5%</td>
<td>97.5%</td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>0.03153*</td>
<td>0.00442</td>
<td>0.00019</td>
<td>0.02264</td>
<td>0.04043</td>
<td></td>
</tr>
<tr>
<td>In(Volume)</td>
<td>-0.51640*</td>
<td>0.09412</td>
<td>0.00445</td>
<td>-0.70200</td>
<td>-0.33480</td>
<td></td>
</tr>
<tr>
<td>Speed-Volume</td>
<td>-0.00004*</td>
<td>0.00001</td>
<td>0.000006</td>
<td>-0.00006</td>
<td>-0.00001</td>
<td></td>
</tr>
<tr>
<td>Curve</td>
<td>0.18200*</td>
<td>0.06565</td>
<td>0.00070</td>
<td>0.05264</td>
<td>0.31110</td>
<td></td>
</tr>
<tr>
<td>Uphill</td>
<td>2.08600*</td>
<td>0.16770</td>
<td>0.00450</td>
<td>1.76600</td>
<td>2.43100</td>
<td></td>
</tr>
<tr>
<td>Downhill</td>
<td>2.82200*</td>
<td>0.16300</td>
<td>0.00453</td>
<td>2.51400</td>
<td>3.15700</td>
<td></td>
</tr>
<tr>
<td>Lanes above2</td>
<td>0.88450*</td>
<td>0.06976</td>
<td>0.00076</td>
<td>0.74870</td>
<td>1.02100</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.52462*</td>
<td>0.54720</td>
<td>0.02643</td>
<td>-5.10200</td>
<td>-2.98800</td>
<td></td>
</tr>
<tr>
<td>ln(Vehicle hours per mile)</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Multiple Vehicle</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>MC error</td>
<td>2.5%</td>
<td>97.5%</td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>-0.02902*</td>
<td>0.00257</td>
<td>0.00012</td>
<td>-0.03380</td>
<td>-0.02375</td>
<td></td>
</tr>
<tr>
<td>In(Volume)</td>
<td>-0.45430*</td>
<td>0.07533</td>
<td>0.00385</td>
<td>-0.59220</td>
<td>-0.30930</td>
<td></td>
</tr>
<tr>
<td>Speed-Volume</td>
<td>0.00005*</td>
<td>0.000009</td>
<td>0.000004</td>
<td>0.00003</td>
<td>0.00006</td>
<td></td>
</tr>
<tr>
<td>Curve</td>
<td>0.49500*</td>
<td>0.04231</td>
<td>0.00054</td>
<td>0.41180</td>
<td>0.57790</td>
<td></td>
</tr>
<tr>
<td>Uphill</td>
<td>2.14800*</td>
<td>0.09754</td>
<td>0.00261</td>
<td>1.96200</td>
<td>2.34500</td>
<td></td>
</tr>
<tr>
<td>Downhill</td>
<td>2.87400*</td>
<td>0.09476</td>
<td>0.00265</td>
<td>2.69400</td>
<td>3.06400</td>
<td></td>
</tr>
<tr>
<td>Lanes above2</td>
<td>1.25500*</td>
<td>0.04675</td>
<td>0.00067</td>
<td>1.16400</td>
<td>1.34600</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.43880*</td>
<td>0.41250</td>
<td>0.02113</td>
<td>-1.80800</td>
<td>-0.22700</td>
<td></td>
</tr>
<tr>
<td>ln(Vehicle hours per mile)</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

*Dbar* 8199.42  *significant at the 95% credible interval*

*pD* 309.239

DIC 8508.66
Table 3: Coefficient estimates of the multivariate model for single vehicle crashes by severity (SV_sev)

<table>
<thead>
<tr>
<th>KSI crashes</th>
<th>Mean</th>
<th>SD</th>
<th>MC error</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed Squared</td>
<td>0.00031*</td>
<td>0.00008</td>
<td>0.000002</td>
<td>0.00016</td>
<td>0.00046</td>
</tr>
<tr>
<td>ln(volume)</td>
<td>-0.80370*</td>
<td>0.07473</td>
<td>0.00206</td>
<td>-0.95020</td>
<td>-0.65740</td>
</tr>
<tr>
<td>Curve</td>
<td>0.26960</td>
<td>0.15170</td>
<td>0.00138</td>
<td>-0.02645</td>
<td>0.56940</td>
</tr>
<tr>
<td>Uphill</td>
<td>1.77200*</td>
<td>0.34860</td>
<td>0.00783</td>
<td>1.12900</td>
<td>2.49500</td>
</tr>
<tr>
<td>Downhill</td>
<td>2.60000*</td>
<td>0.33390</td>
<td>0.00778</td>
<td>1.99000</td>
<td>3.30400</td>
</tr>
<tr>
<td>Lanes above2</td>
<td>0.82470*</td>
<td>0.16360</td>
<td>0.00169</td>
<td>0.50910</td>
<td>1.15000</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.37560*</td>
<td>0.60270</td>
<td>0.02098</td>
<td>-5.08800</td>
<td>-2.74800</td>
</tr>
<tr>
<td>ln(Vehicle hours per mile)</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SL crashes</th>
<th>Mean</th>
<th>SD</th>
<th>MC error</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed Squared</td>
<td>0.00019*</td>
<td>0.00003</td>
<td>0.00001</td>
<td>0.00013</td>
<td>0.00026</td>
</tr>
<tr>
<td>ln(volume)</td>
<td>-0.75780*</td>
<td>0.03620</td>
<td>0.00102</td>
<td>-0.82770</td>
<td>-0.68650</td>
</tr>
<tr>
<td>Curve</td>
<td>0.17010*</td>
<td>0.07105</td>
<td>0.00069</td>
<td>0.03053</td>
<td>0.30880</td>
</tr>
<tr>
<td>Uphill</td>
<td>2.10300*</td>
<td>0.18000</td>
<td>0.00450</td>
<td>1.76300</td>
<td>2.46900</td>
</tr>
<tr>
<td>Downhill</td>
<td>2.82500*</td>
<td>0.17500</td>
<td>0.00451</td>
<td>2.49300</td>
<td>3.17900</td>
</tr>
<tr>
<td>Lanes above2</td>
<td>0.89620*</td>
<td>0.07684</td>
<td>0.00082</td>
<td>0.74730</td>
<td>1.04900</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.73460*</td>
<td>0.29220</td>
<td>0.01012</td>
<td>-2.82800</td>
<td>-1.68700</td>
</tr>
<tr>
<td>ln(Vehicle hours per mile)</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Dbar* 4203.77 *significant at the 95% credible interval
*pD* 131.66
*DIC* 4335.43
Table 4: Coefficient estimates of the multivariate model for multiple vehicle crashes by severity (MV_ref).

<table>
<thead>
<tr>
<th>KSI crashes</th>
<th>Mean</th>
<th>SD</th>
<th>MC error</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>0.030577*</td>
<td>0.006709</td>
<td>0.000277</td>
<td>0.017187</td>
<td>0.044014</td>
</tr>
<tr>
<td>Volume Squared</td>
<td>0.00001*</td>
<td>0.02502</td>
<td>0.00090</td>
<td>0.03514</td>
<td>0.13260</td>
</tr>
<tr>
<td>Speed-Velocity</td>
<td>-0.00011*</td>
<td>0.00002</td>
<td>0.000001</td>
<td>-0.00015</td>
<td>-0.00008</td>
</tr>
<tr>
<td>Curve</td>
<td>0.43220*</td>
<td>0.09468</td>
<td>0.00100</td>
<td>0.24760</td>
<td>0.61980</td>
</tr>
<tr>
<td>Uphill</td>
<td>1.96400*</td>
<td>0.24250</td>
<td>0.00622</td>
<td>1.50500</td>
<td>2.45900</td>
</tr>
<tr>
<td>Downhill</td>
<td>2.85200*</td>
<td>0.23370</td>
<td>0.00625</td>
<td>2.41300</td>
<td>3.33000</td>
</tr>
<tr>
<td>Lanes above2</td>
<td>1.25300*</td>
<td>0.11030</td>
<td>0.00120</td>
<td>1.03900</td>
<td>1.47300</td>
</tr>
<tr>
<td>Intercept</td>
<td>-7.6586*</td>
<td>0.44410</td>
<td>0.01844</td>
<td>-8.07400</td>
<td>-6.32900</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ln(Vehicle hours per mile)</th>
<th>Mean</th>
<th>SD</th>
<th>MC error</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL crashes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>-0.01132*</td>
<td>0.00374</td>
<td>0.00017</td>
<td>-0.01839</td>
<td>-0.00413</td>
</tr>
<tr>
<td>Volume Squared</td>
<td>0.000004*</td>
<td>0.00001</td>
<td>0.0000006</td>
<td>0.000002</td>
<td>0.000007</td>
</tr>
<tr>
<td>Speed-Velocity</td>
<td>-0.00003*</td>
<td>0.00001</td>
<td>0.0000004</td>
<td>-0.00005</td>
<td>-0.00004</td>
</tr>
<tr>
<td>Curve</td>
<td>0.48210*</td>
<td>0.04645</td>
<td>0.00062</td>
<td>0.39050</td>
<td>0.57290</td>
</tr>
<tr>
<td>Uphill</td>
<td>2.14900*</td>
<td>0.10250</td>
<td>0.00276</td>
<td>1.95500</td>
<td>2.35600</td>
</tr>
<tr>
<td>Downhill</td>
<td>2.85300*</td>
<td>0.09973</td>
<td>0.00275</td>
<td>2.66400</td>
<td>3.05600</td>
</tr>
<tr>
<td>Lanes above2</td>
<td>1.21600*</td>
<td>0.05064</td>
<td>0.00078</td>
<td>1.11700</td>
<td>1.31600</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.2386*</td>
<td>0.2322</td>
<td>0.01048</td>
<td>-4.197</td>
<td>-3.292</td>
</tr>
</tbody>
</table>

| ln(Vehicle hours per mile)  | 1    |     |         | -     | -     |

<table>
<thead>
<tr>
<th>Dbar</th>
<th>6588.38</th>
<th></th>
<th>*significant at the 95% credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>pD</td>
<td>266.504</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>6854.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5: Elasticity of speed and the minimum (speed increases by 2.5 mph) and maximum (speed increases by 5 mph) expected increase of crashes by type.

<table>
<thead>
<tr>
<th>Crash Type</th>
<th>Elasticity</th>
<th>Percentage of expected crash increase</th>
<th>Expected additional crashes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Min.</td>
<td>Max.</td>
</tr>
<tr>
<td>SV KSI</td>
<td>2.595</td>
<td>10.017</td>
<td>20.059</td>
</tr>
<tr>
<td>SV SL</td>
<td>1.590</td>
<td>6.137</td>
<td>12.291</td>
</tr>
<tr>
<td>MV KSI*</td>
<td>0.886</td>
<td>3.420</td>
<td>6.849</td>
</tr>
</tbody>
</table>

*Estimation based on the average volume conditions (i.e. 148 vehicles per lane)
Figure 1: Crashes per vehicle hours travelled per mile as a function of a) speed for all single vehicle crashes, b) volume for all single vehicle crashes, c) speed for all multiple vehicle crashes and d) volume for all multiple vehicle crashes.
Figure 2: 3D contour plot of crash rate for all Single Vehicle crashes as a function of the speed and volume conditions.

Figure 3: 3D contour plot of crash rate for all Multiple Vehicle crashes as a function of the speed and volume conditions.
Figure 4: Crashes per vehicle hours travelled per mile as a function of a) speed for KSI single vehicle crashes, b) volume KSI single vehicle crashes, c) speed for SL single vehicle crashes and d) volume for SL single vehicle crashes.
Figure 5: 3D contour plot of crash rate for KSI Single Vehicle crashes as a function of the speed and volume conditions.

Figure 6: 3D contour plot of crash rate for SL Single Vehicle crashes as a function of the speed and volume conditions.
Figure 7: Crashes per vehicle hours travelled per mile as a function of a) speed for KSI multiple vehicle crashes, b) volume for KSI multiple vehicle crashes, c) speed for SL multiple vehicle crashes and d) volume for SL multiple vehicle crashes.
Figure 8: 3D contour plot of crash rate for KSI Multiple Vehicle crashes as a function of the speed and volume conditions.

Figure 9: 3D contour plot of crash rate for SL Multiple Vehicle crashes as a function of the speed and volume conditions.