“I did not see that coming”: A latent variable structural equation model for understanding the effect of road predictability on crashes along horizontal curves

Afghari, Amir Pooyan; Vos, Johan; Farah, Haneen; Papadimitriou, Eleonora

DOI
10.1016/j.aap.2023.107075

Publication date
2023

Document Version
Final published version

Published in
Accident Analysis and Prevention

Accident Analysis and Prevention 187 (2023) 107075

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

“I did not see that coming”: A latent variable structural equation model for understanding the effect of road predictability on crashes along horizontal curves

Amir Pooyan Afghari [a,*], Johan Vos [b], Haneen Farah [b], Eleonora Papadimitriou [a]

[a] Safety and Security Science Section, Faculty of Technology, Policy and Management, Delft University of Technology, Delft, Netherlands
[b] Transport and Planning Department, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands

ARTICLE INFO

Keywords:
Anticipation
Predictability
Crash risk
Horizontal curve
Structural equation modelling

ABSTRACT

Driver anticipation plays a crucial role in crashes along horizontal curves. Anticipation is related to road predictability and can be influenced by roadway geometric design. Therefore, it is essential to understand which geometric design elements can influence anticipation and cause the road to be (un)predictable. This exercise, however, is not straightforward because anticipation is individual-specific whereas road geometric design is location-specific; anticipation is latent and measuring it may not be trivial; anticipation may have several stages from the preceding tangent until the midst of the curve; and not all drivers anticipate in the same way and thus there may well be unobserved heterogeneity in the effect of anticipation on crash risk.

Despite methodological advancements in crash risk modelling, there is no econometric model that can adequately explain the above complexities. This study aims to fill this gap by developing an econometric model with a new latent variable, named ‘predictability’ that is measured by individual-specific driving behaviour indicators and predicted by location-specific road geometric factors. The model is specified with random parameters to account for unobserved heterogeneity and is empirically tested by a unique dataset including detailed geometric design and driver behaviour data obtained for 156 curves in the Netherlands. Results indicate that higher exposure and uphill vertical grade are associated with increased likelihood of vehicle crashes along horizontal curves, whereas adequate superelevation and higher predictability are associated with decreased likelihood of those crashes. Pavement friction influences this likelihood too but it has varied effects. Road predictability is influenced by the differences in angle of horizontal curves, vertical grades, and width of consecutive road segments.

1. Background

Horizontal curves are among the roadway locations with high prevalence of vehicle crashes across the world. In the Netherlands, 4 out of 62 fatal crashes that were recorded on national roads in 2019 were directly related to sharp curves (Davidse et al., 2020). A detailed investigation of those crashes revealed that the road infrastructure at the crash locations was not predictable and so the drivers involved in those crashes did not expect the upcoming sharp curves (Davidse et al., 2020). Such lack of ‘predictability’ caused two of the drivers running off the road and falling into an adjacent ditch while the other two drivers hit a tree with deadly consequences. Similar findings have been reported elsewhere in the world. Findings from the 100-Car Naturalistic Driving Study in the United States showed that 30% of vehicle crashes and near-misses occurred along curves (McLaughlin et al., 2009) and 6% of fatal crashes were directly related to sharp curves. These statistics raise the flag for horizontal curves as locations with potential high risk of vehicle crashes that need to be treated by highway engineers and road safety experts. It is, therefore, essential to understand what factors contribute to these crashes along horizontal curves.

Previous studies have shown that factors contributing to vehicle crashes may arise from distinct sources of risk, such as human factors or driver behaviour, road geometry and spatial features of the roadside environment (Afghari et al., 2018; Shaon et al., 2019). One of the most

* Corresponding author.
E-mail addresses: a.p.afghari-1@tudelft.nl (A.P. Afghari), j.vos-1@tudelft.nl (J. Vos), h.farah@tudelft.nl (H. Farah), e.papadimitriou@tudelft.nl (E. Papadimitriou).

https://doi.org/10.1016/j.aap.2023.107075
Received 8 November 2022; Received in revised form 4 April 2023; Accepted 12 April 2023
Available online 21 April 2023

common human factors contributing to crashes is situation awareness which is defined by Endsley (2017) as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". The ability to project the status of different elements in the environment (level 3 in Endsley model), is also known as anticipation. Anticipation is a key element in driving along curves; it guides drivers to adjust their speed and lane position to accommodate the severity of the curve (Reymond et al., 2001). Several studies provide an important insight about driver's anticipation along horizontal curves: it is entangled with roadway geometric design. This entanglement has been referred to as self-explaining road design (Theeuwes, 2021; Theeuwes & Godthelp, 1995; Walker et al., 2013) or geometric design consistency (D. Llopis-Castelló et al., 2018). A self-explaining/consistent road induces driver expectations and elicits appropriate driving behaviour automatically (Kuiken et al., 2012). Therefore, it is very important to understand which geometric design elements may influence anticipation and cause the roadway to be (un)predictable.

One way of answering the above question is by developing an econometric model of crashes and including there-in an independent variable capturing the interaction of anticipation and roadway geometric design. However, this exercise is not straightforward due to the following challenges: *(i)* anticipation is related to drivers and is individual-specific whereas geometric design is related to roads and is location-specific; *(ii)* anticipation is usually not observed but *latent* among drivers and therefore, measuring it may not be straightforward; *(iii)* drivers' anticipation along horizontal curves may have several stages starting from the preceding tangent but lasting until the midst of the curve itself; and *(iv)* not all drivers anticipate in the same way and thus there may well be *unobserved heterogeneity* in the effects of anticipation on crash risk.

Despite methodological advancements in crash risk modelling, there is no econometric model that can adequately explain the interaction between driver anticipation and roadway geometry and quantify its effect on crash risk along horizontal curves. The objective of this research is to fill this gap by developing a new econometric model that can adequately explain the interaction between driver anticipation and roadway geometry and quantify its effect on crash risk along horizontal curves, while at the same time addressing the above conceptual and methodological challenges: latent nature of anticipation and unobserved heterogeneity. The proposed econometric model includes a new latent variable that is measured by individual-specific driving behaviour indicators and predicted by location-specific road geometric factors. The model is specified with random parameters to account for unobserved heterogeneity.

The remainder of this paper is organised as follows: Section 2 presents a brief review of the state-of-the-art on the impact of road geometric design features and anticipation on crashes, with a focus on the existing modelling techniques. Section 3 presents the proposed methodological approach and model formulation in this study, and Section 4 presents the description of the dataset. Section 5 presents the modelling results and provides further discussion about those results. Section 6 gives the conclusions of this paper.

2. Literature review

According the American Association of State Highway and Transportation Officials, there are four key features in designing horizontal curves: radius, design/operating speed, side friction factor, and superelevation (AASHTO, 2001). It has been well established by *meta*-analysis that low curve radius, especially lower than 200 m, is associated with increased crash risk (Elvik, 2013). The negative effect of low curve radius is deteriorated when horizontal curves overlap with sharp vertical grades (crest or sag) (Bauer & Harwood, 2013), where often adequate sight distance is not available (Papadimitriou et al., 2018). Poor pavement conditions or rainy weather may increase crash risk on horizontal curves too (Donnell et al., 2016). Moreover, the absence of transition curves (clothoids) is associated with negative safety outcomes (Martensen et al., 2019).

In addition to the above design features, human factors related to drivers perceptions and expectations, as well as comprehension, decision and task execution capabilities are important parameters to be taken into account in curve design (Campbell et al., 2008). Especially with respect to the anticipation (as defined in the previous section), design consistency has been found to have a positive safety effect e.g. Montella and Imbriani (2015); for instance, in terms of road bendiness (i.e. number of consequent curves per road length), drivers may have better anticipation and thus better speeding behaviour over subsequent curves rather than in isolated ones (Van Petegem & Schermers, 2016). Anticipation precedes curve entry by about 4 s (Campbell, 2012) and is mostly based on a top-down process, i.e., schema where drivers' expectations depend on their past experiences (Borsos et al., 2015). Within this schema, however, the perceptual cues in the environment trigger the activation of a certain behaviour (in this case the desired speed) (Charlton & Starkey, 2017). Findings from previous studies (Charlton, 2007; Lewis-Evans & Charlton, 2006) support this notion and suggest while advance warning signs (even with advisory speed signs) are not an effective measure to make drivers slow down, chevrons or curve warnings that highlight or delineate the sharpness of the curve are effective in substantially reducing drivers' speed along curves. These curve warnings provide perceptual cues that can be processed "bottom-up". Nama et al. (2020) found that curve radius and the length of the preceding tangent are the most influential factors affecting driver behaviour while approaching a curve. Finally, Lehtonen et al. (2012) observed that drivers anticipate open curves by switching their visual attention between the road and the occlusion point.

Alternative methodologies exist in the transport safety literature to create an econometric model for the inter-relationship between road geometry, driver anticipation and crashes on curves. The most straightforward approach is to include a control variable in the econometric model to directly measure the interaction between anticipation and road geometry. Azmeri Khan et al. (2023) developed a negative binomial Lindley model for run-off road crashes and included geometric design consistency as the control variable. They found that any sudden change in roadway geometry (the control variable) increases the risk of run-off road crashes. To capture the unobserved heterogeneity in the data (Mannering et al., 2016), they specified their model with random parameters. Afghari et al. (2018) proposed an instrumental variable model of crashes in a joint econometric framework to account for the inter-relationship between risk factors and empirically showed that the proposed approach provides more insight into those interactions. A few other studies used instrumental variable modelling as well to investigate the inter-relationship between sleepiness and headway (Afghari et al., 2022), speed enforcement and safety (Yasmin et al., 2022). They used a random parameters specification too for addressing unobserved heterogeneity. However, neither of the two approaches are able to address the latent nature of anticipation and its several stages along curves. More specifically, the unobserved part of the interaction between anticipation and road geometry (due to measurement error or lack of data) may not be captured via the direct control variable or the instrumental variable. Therefore, any inference that is made about such interaction using either of the two approaches, may not be totally accurate.

An alternative approach to address this issue is by creating a latent variable for such an interaction in a structural equation modelling (SEM) framework and including it, as well as its error component, into the overall econometric model. SEM is an umbrella for a wide range of methods, from simple confirmatory factor analysis (Brown & Moore, 2012) to complex likelihood-based latent variable models (Afghari et al., 2019; Paschalidis et al., 2019), with the latter providing unique capabilities for addressing unobserved heterogeneity. However, the potential benefit of latent variable modelling for capturing the interaction

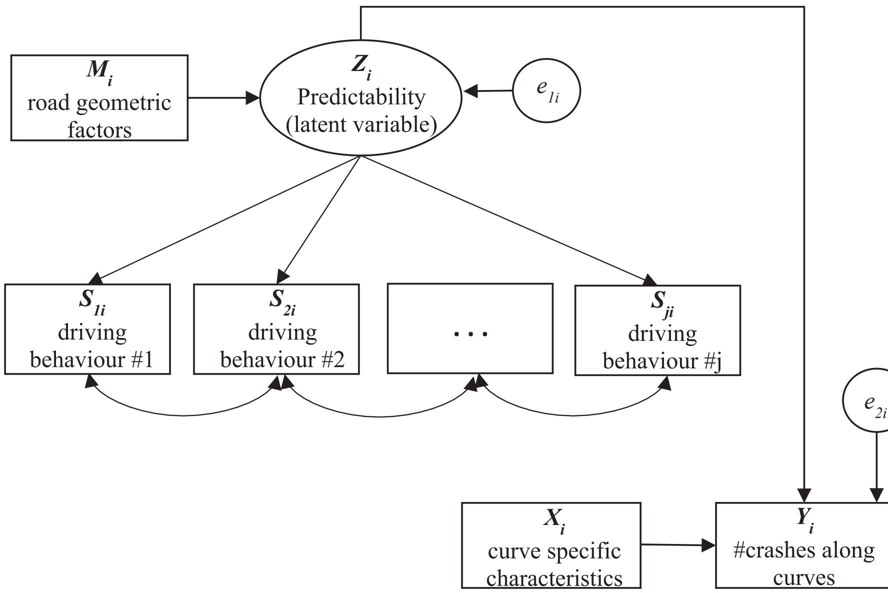

Fig. 1. Path diagram corresponding with the theoretical construct behind the relationship between predictability, road geometric factors and crash risk.

between drivers' anticipation and road geometry has been unexplored.

3. Methods

To overcome the primary challenge in modelling anticipation and roadway geometric design (their individual- versus location-specific nature), it is hypothesized that these two factors are linked via a new latent variable which is labelled as 'predictability' in this manuscript. This latent variable is measured by individual-specific driving behaviour indicators (e.g. drivers' speed and acceleration) and predicted by location-specific geometric factors (e.g. difference in radii of horizontal curves). Prior to developing a model for such a latent variable, a theoretical construct should be established for the relationships between all variables (observed and latent). This theoretical construct is usually illustrated by *path diagrams* in which latent variables are shown by ellipses, observed variables by rectangles, unobserved error terms by small circles, and the relationships between variables by directional arrows (Hox & Bechger, 1998).

3.1. Theoretical construct and path diagram

The path diagram corresponding with the theoretical construct in this study is shown in Fig. 1 below. It is hypothesised that the number of vehicle crashes along curves ($\boldsymbol{Y_i}$) is an observed indicator of crash risk, and the factors contributing to this indicator arise from observed curve-specific characteristics ($\boldsymbol{X_i}$) and a latent variable labelled as *predictability* ($\boldsymbol{Z_i}$). Moreover, the latent predictability can be influenced by roadway geometric factors ($\boldsymbol{M_i}$) preceding the curve as well as along the curve. Finally, it is assumed that the latent predictability can be measured by several indicators of driving behaviour ($\boldsymbol{S_{ij}}$) such as speed, deceleration, and the distance at which drivers start to change their behaviour prior to the curve. This overall hypothesis will be tested using an advanced latent variable econometric model (discussed later) estimated against empirical data.

3.2. Principal component analysis

It is hypothesized that anticipation has several stages starting prior to and lasting until the midst of the curve, and thus predictability as a latent variable may need to be measured through several driving behaviour indicators at different stages too, each of which may capture certain stage of predictability. For example, driving speed prior to the curve may capture the first stages of predictability whereas speed oscillations along the curve may capture later stages of predictability. However, driver behaviour underlies all of these indicators and thus they may have high autocorrelation with one another (bi-directional arrows in the above path diagram) as previously shown in the statistical analysis of behavioural data (Huitema & McKean, 1991). Principal Component Analysis (PCA) is a common approach used in the statistical (Tipping & Bishop, 1999) and transport (Afghari et al., 2021a, Afghari et al., 2021b) literature to summarize data when there are too many variables interacting with one another in the analysis. Hence, it is used in this study too with the purpose of including as many indicators as possible for predictability while minimizing the autocorrelations between those indicators.

The PCA creates a set of new variables, referred to as principal components (PC), which are orthogonal to one another, and each component is a linear combination of the original set of variables. The principal components can be obtained by applying the orthogonal transformation and finding the Eigenvectors and Eigenvalues of the Spearman correlation matrix of the original set of explanatory variables. The principal components are then ranked based on their decreasing contribution to the total variance of the original set of explanatory variables: the first principal component explains the highest variability in the explanatory variables; the second one explains the second highest variability and so forth (the cumulative contribution of all principal components is equal to 1). These principal components can then be used in the analysis as representatives of the original set of variables. The number of components to be used in the model depends on the specific research objective, though the optimum number of components is the number of components with Eigenvalues greater than one (Tipping & Bishop, 1999). In the context of this study, the first principal component is selected as the most appropriate representative of all indicators for road's predictability.

3.3. Latent variable random parameters negative binomial model

The hypothesized path diagram is now quantified using an advanced econometric model. Let Y_i represent the number of crashes along the i^{th} curve in a road network. Vehicle crashes are the outcomes of a Poisson process (Lord et al., 2005) and so Y_i follows a Poisson distribution with mean μ_i:

$$Y_i \ Poisson(\mu_i) \tag{1}$$

Assuming an exponential function for the mean of the Poisson distribution, the expected number of crashes (μ_i) along the i^{th} curve can be expressed as a function of explanatory variables (Afghari et al., 2018):

$$\mu_i = exp(\beta_i X_i)\exp(\varepsilon_i) \quad (2)$$

Where β_i are estimable parameters (including the intercept), X_i are explanatory variables, and $\exp(\varepsilon_i)$ is a random error term following a Gamma distribution with mean 1 and variance $1/\varphi$. To account for unobserved heterogeneity in the data, model parameters (β_i) are allowed to vary across curves. Such a model specification is referred to as *random parameters negative binomial model* (Anastasopoulos & Mannering, 2009) in which the parameters are assumed to follow probabilistic distributions (e.g. normal, uniform, triangular, etc.) across observations. The probability density function of this model can be obtained by:

$$P(Y_i = y_i|\beta_i, \varphi) = \int \frac{\Gamma(\varphi + y_i)}{\Gamma(\varphi)y_i!}\left(\frac{\varphi}{\varphi + \mu_i}\right)^{\varphi}\left(\frac{\mu_i}{\varphi + \mu_i}\right)^{y_i} f(\beta_i)d\beta_i \quad (3)$$

where $\Gamma(\cdot)$ is the gamma function and $f(\beta_i)$ is the density of the model parameters. Predictability is now incorporated into the model as a latent variable. More specifically, we define a new latent variable (Z_i) representing the predictability, and insert it with a parameter (α_i) into the mean function of the random parameters negative binomial model:

Table 1
Summary statistics of road geometric characteristics and driving behaviour indicators for the road network in this study.

Variable	Minimum	Maximum	Mean	SD
Curve characteristics				
Number of crashes	0	27	1.82	4.05
Length [m]	31	1018	298	210
AADT [vehicles]	800	64,100	18,274	13,312
Horizontal Radius [m]	60	801	297	270
Deflection angle [grad.]	5.55	283.74	85.38	68.3
Vertical grade [%]	−3	+3	0	1
Number of lanes	1	4	1.64	0.76
Road width [m]	5.02	22.18	10.9	3.66
Superelevation [%]	1	6	4	1
Speed limit [km/h]	50	100	95	12
Minimum measured friction coefficient	0.38	0.79	0.58	0.08
Stopping sight distance at curve start [m]	106	971	385	127
Characteristics of the preceding road segment				
Horizontal Radius [m]	128	∞*	669,736	470,808
Deflection angle [grad.]	0	183	5	23
Vertical grade [%]	−2	3	0	1
Number of lanes	1	6	2.46	1.08
Road width [m]	6.84	43.55	14.36	5.46
Superelevation [%]	0	6	3	1
Stopping sight distance 100 m before curve [m]	66	1181	310	161
Driving behaviour indicators				
85th percentile speed at curve start [km/h]	64	138	103	18
85th percentile speed at end of deceleration [km/h]	60	137	101	20
Standard deviation of deceleration at maximum [m/s^2] deceleration point	0.19	1.84	0.48	0.19
Median distance between start point of deceleration and curve start [m]	15	531	192	132
Median distance between maximum deceleration point and curve start [m]	0	278	33	37
Median distance between curve start and end of Deceleration [m]	14	412	95	73

*indicating a straight segment.

$$\mu_i = exp(\beta_i X_i + \alpha_i Z_i)\exp(\varepsilon_i) \quad (4)$$

It is hypothesized that this latent variable can be measured by a linear combination of driving behaviour indicators obtained from the principal component analysis. As such, a measurement equation is defined for this latent variable as:

$$PC_1 = h(\gamma Z_i) \text{and} PC_1 = \sum_{j=1}^{J} w_j S_{ji} \quad (5)$$

where PC_1 is the first principal component of driving behaviour indicators (S_{ji}), w_j are factor loadings of the driving behaviour indicators within the first principal component, $h(.)$ is the standard normal distribution probability density function, Z_i is the latent variable, and γ is an estimable parameter. Meanwhile, a separate structural equation is also defined for the latent variable to correlate it with a number of exogenous explanatory variables:

$$Z_i = \lambda m_i + \xi_i \quad (6)$$

where λ are estimable parameters, m_i are exogenous explanatory variables, and ξ_i is a normally distributed error term with mean 0 and variance σ^2. The probability density function of the overall model can then be obtained by:

$$P(Y_i = y_i|\beta_i, \varphi, \alpha, \gamma, \lambda, \sigma) = \iint \frac{\Gamma(\varphi + y_i)}{\Gamma(\varphi)y_i!}\left(\frac{\varphi}{\varphi + \mu_i}\right)^{\varphi}\left(\frac{\mu_i}{\varphi + \mu_i}\right)^{y_i} f(\beta_i)g(Z_i)h(\gamma Z_i)d\beta_i dZ_i \quad (7)$$

where $g(Z_i)$ is the density of the latent variable (Equation (6) and the rest of notations are as previously stated. The likelihood function (L) of the overall model can be obtained by the product of the above density function over the entire observations:

$$L = \prod_{i=1}^{N} P(Y_i = y_i|\beta_i, \varphi, \alpha, \gamma, \lambda, \sigma) \quad (8)$$

This elaborate model is referred to as *latent variable random parameters negative binomial (LV-RPNB)* in this manuscript and does not have a closed form to be estimated using regular maximum likelihood estimation technique. Therefore, maximum simulated likelihood estimation is used where quasi random draws from Halton sequences are employed to simulate the densities of the random parameters and the latent variable (Bhat, 2001). It has been shown that this simulated maximum likelihood estimator is unbiased and consistent for a large number of draws (Munkin & Trivedi, 1999).

3.4. Model selection

To test the performance of the LV-RPNB model, its statistical fit to empirical data is evaluated and compared with that of the regular RPNB model without latent predictability. While Akaike Information Criterion and Bayesian Information Criterion are widely employed to compare the statistical fit of models that have the same likelihood structure, these measures of fit are not comparable between the models that have different likelihood structures (with versus without latent variable). As such, Mean Absolute Deviance (MAD) and Mean Squared Predictive Error (MSPE) are used to compare the performance of the models in this study. Suppose Y_i and $\widehat{Y}_i$ are the observed and the predicted values, respectively, of the dependent variable for curve i. The MAD and MSPE are calculated as (Washington et al., 2020):

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|Y_i - \widehat{Y}_i| \quad (9)$$

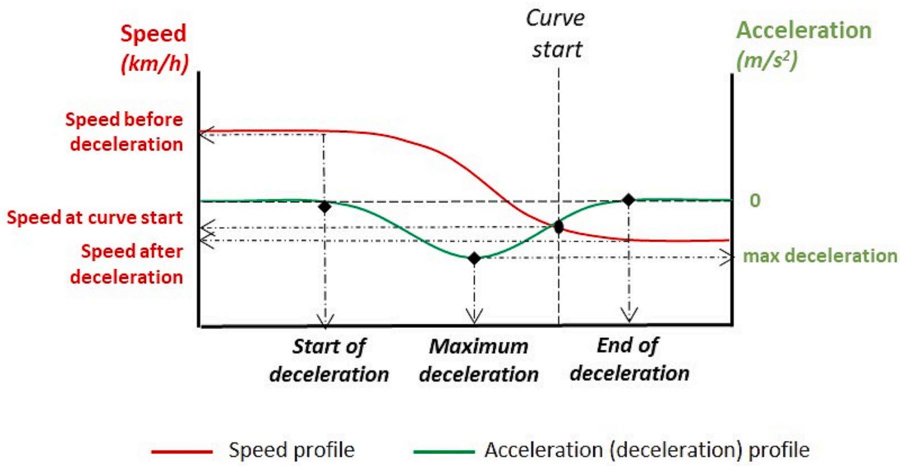

Fig. 2. Theoretical speed and acceleration (deceleration) profiles for the curve approach.

$$MSPE = \frac{1}{N-P}\sum_{i=1}^{N}(Y_i - \widehat{Y}_i)^2 \quad (10)$$

where N is the sample size and P is the number of estimated parameters. The model with smaller MAD and MSPE is usually preferred over the other model.

The above measures of fit are based on model predictions and thus may not properly reflect the complexity of the models. As such, statistical fit of the models is also assessed using McFadden pseudo-rho squared – adjusted (ρ^2_{adj}):

$$\rho^2_{adj} = 1 - \left[\frac{LL_{Full} - P}{LL_0}\right] \quad (11)$$

where LL_{Full} and LL_0 are the log-likelihoods of the full and the null models, respectively and the rest of notations are as previously stated. ρ^2_{adj} is analogous to adjusted R^2_{adj} in linear models and so a higher ρ^2_{adj} indicates improved statistical fit.

4. Empirical data

Empirical data for this study were extracted from a larger dataset used by Vos et al. (2021b), which contained detailed road geometric information for 156 horizontal curves in the Netherlands. Of these curves, 99 are first curves (preceded by a tangent and are therefore not influenced by a preceding curve), 47 are isolated curves (preceded and followed by a tangent), and 41 are reverse curves (followed immediately by a curve in the opposite direction). The geometric data include length of curves, horizontal radius, deflection angle, number of lanes, road width, distances from the edge marking to the barrier on the left and right side, presence of emergency lane, width of emergency lane, sight distance, super elevation, vertical grade, speed limit, and stated advisory speed. The data were further enriched by adding road friction and traffic volume obtained from Rijkswaterstaat –the Ministry of Infrastructure and Water Management in the Netherlands. Friction data were collected in 2020 and consisted of skid resistance at each hectometre, reported as a friction coefficient ranging from 0 to 1 (Vos et al., 2017). These measurements were added to the dataset in this study as the minimum, maximum and average friction per curve. Traffic data were collected in 2019 and consisted of Annual Average Daily Traffic (AADT) per curve. In addition, three years of vehicle crashes (from 2018 to 2020) were obtained from the VIA Traffic Solutions Software (VIA, 2020) and were assigned to the curves in this study based on their geographic coordinates (latitude and longitude). The summary statistics of the road geometric characteristics of the curves is shown in Table 1.

4.1. Driving behaviour indicators

In addition to the road geometric data, 996,375 individual free-flow speed profiles were also collected along the curves and were added to the data in this study. The speed characteristics for each curve were determined using High Frequency Floating Car Data (HF FCD) with the data collection frequency of 1 Hz along the selected freeway sections. The data were extracted from "Flitsmeister" mobile phone application which is used by approximately 1.6 million users in the Netherlands –roughly 15% of all driver-licence holders. During the months of March, April and September of 2020, over 12 million unique speed profiles were obtained. The data were cleaned out of the periods where road works were present, and the periods in which more than 5 vehicles per lane per minute were present, in order to obtain free-flow speed profiles only.

For each individual speed profile, a deceleration profile was derived. Both profiles were related to the position of the start of the curve, as is shown in Fig. 2. The following variables were derived per profile: (i) distance from start of deceleration to curve start, and speed before deceleration; (ii) distance to curve start where maximum deceleration was reached, including the amount of maximum deceleration reached; (iii) speed at curve start; and (iv) distance from curve start and end of deceleration, and the speed at that position. For each curve, the median position of start of deceleration, maximum deceleration and end of deceleration were calculated. Furthermore, for each curve the 85th percentile of speed before the curve, maximum deceleration, speed at curve start and speed in the curve were calculated.

5. Results and discussion

The RPNB and LV-RPNB models were estimated against the above empirical data and their performances were compared to assess the suitability of the latent variable modelling approach. In both of the models, explanatory variables were selected using a stepwise variable selection criterion. Explanatory variables were tested for multicollinearity by computing the Pearson or Spearman correlation coefficients, and the variables with unacceptably high (>0.7) correlation coefficients were not simultaneously introduced into the model. In addition, the principal component analysis was applied on several driving behaviour indicators within the LV-RPNB model. The parameters of all variables were tested for random parameters specification and normal distribution was used as the distribution for all of the random parameters. The parameters were considered random only if their standard deviations are statistically significant. The models were estimated using the maximum simulated likelihood approach with 2000 Halton draws. The required number of Halton draws was selected so that further increasing the number of draws does not change the estimates

Table 2
Results of principal component analysis (PCA) of driving behaviour indicators.

		1st Principal component	2nd Principal component	3rd Principal component	4th Principal component	5th Principal component
Factor loadings	Median distance between start point of deceleration and curve start	0.683	0.188	−0.060	−0.300	−0.636
	Median distance between maximum deceleration point and curve start	0.044	−0.741	−0.160	0.513	−0.400
	Standard deviation of deceleration at maximum deceleration point	−0.355	−0.198	−0.712	−0.562	−0.107
	Difference between 85th percentile speed at curve start and constant speed point	−0.164	−0.468	0.663	−0.548	−0.119
	Median distance between curve start and constant speed point	0.616	−0.395	−0.156	−0.174	0.640
Eigenvalues		1.816	1.438	1.020	0.620	0.106
Proportion of variance explained by the principal component		0.363	0.288	0.204	0.124	0.021
Cumulative proportion of variance explained by the principal component		0.363	0.651	0.855	0.979	1.000

Table 3
Regression results of random parameters negative binomial (RPNB) model.

	Estimate	SE	t	p value
Constant	−14.103	2.806	−5.027	<0.001
Log(AADT)	1.661	0.163	10.194	<0.001
Standard deviation Log(AADT)	0.001	0.306	0.004	0.997
Deflection angle of horizontal curve	0.008	0.002	3.821	<0.001
Superelevation >2%	−1.559	0.490	−3.181	0.001
Positive vertical grade	0.560	0.288	1.949	0.051
Minimum friction	−3.461	1.799	−1.924	0.054
Standard deviation of minimum friction	1.938	0.244	7.929	<0.001
Dispersion parameter	28.365	3.099	9.136	<0.001

significantly. While estimating the LV-RPNB model, the dispersion parameter of the negative binomial distribution was fixed for the purpose of identification.

5.1. Indicators of predictability

Prior to estimating the models, several driving behaviour variables were selected as observed indicators of predictability. These variables and the rationale behind selecting them are discussed in the following: *(a)* The median distance between the starting point of deceleration and the start of the curve: this variable was selected to represent early stages of predictability at which drivers first notice the changes in driving task demand and start reacting to such changes. A such, higher distance between these two points may indicate higher predictability; *(b)* The median distance between the maximum deceleration point and the start of the curve: this variable was selected to represent middle stages of predictability at which drivers apply maximum deceleration and prepare for entering the curve; as such, higher distance between these two points may indicate higher predictability too; *(c)* The standard deviation of deceleration at the maximum deceleration point prior to the curve: this variable was selected to represent middle stages of predictability too but also to reflect the variance of predictability among drivers. Higher standard deviations of deceleration at this point may indicate lower predictability of the road among drivers; *(d)* The difference between 85th percentile of speed at the start of the curve and at the point where speed is constant (end of deceleration) along the curve: this variable was selected to represent late stages of predictability at which drivers are still adapting to a comfortable speed along the curve. As such, higher differences in the speeds may indicate lower predictability of the road; and *(e)* The median distance between the start of the curve and the point where speed is constant along the curve: this variable was selected to represent late stages of predictability as well. Higher median distance between these two points may indicate higher predictability.

The above variables were summarized into five orthogonal principal components which if combined explain the full observed variability in predictability. The results of the principal component analysis, the eigenvalues and the proportion of explained variability are shown in Table 2. The factor loadings of variables within the first principal component (showing the direction of their association with this component) are in line with the abovementioned rationale for the selected variables. As such, the first principal component may be used as an indicator of predictability in the measurement equation of the LV-RPNB model.

5.2. Model estimation results

Estimation results of the regular RPNB model presented in Table 3 show that of all variables, AADT, deflection angle of horizontal curves, adequate superelevation, positive vertical grade, and minimum friction are statistically significant with 5% significance level.

The positive parameter of AADT (1.661) indicates that higher exposure is associated with increased likelihood of vehicle crashes along curves, which is according to expectations and previous literature (AASHTO, 2001). Similarly, the positive parameter of deflection angle of horizontal curves (0.008) indicates that larger angles (sharper curvature) are associated with increased likelihood of crashes. Similar findings have been reported in the literature (Ma et al., 2020; Schneider et al., 2010; Xin et al., 2019) implying adverse effects of horizontal curves on crash risk. On the contrary, the negative parameter of categorical superelevation (-1.559) indicates that a superelevation higher than 2% decreases the likelihood of crashes along curves. This finding is in line with the previous findings about the effects of adequate superelevation on crashes (Papadimitriou et al., 2019; Peng et al., 2021). The parameter of positive vertical grade (0.560) indicates that crashes are more likely on uphill curves, in comparison with downhill curves or curves without a vertical grade. This finding might reflect the lack of sufficient sight distance when horizontal curves are combined with upward vertical grades. Chang (2005) found that sections with severe uphill/downhill grade (3% or more) have higher likelihood of crash occurrence compared to level sections. In this study, however, the effect of downhill curve on crash risk was not found to be statistically significant. This finding might be due to very low ranges of the downhill grade (not exceeding −3%) in the study area. In other words, the downhill grade in our dataset is relatively mild and does not significantly impact the driving speed. The literature with respect to the 3D alignment parameters, including the combinations of the horizontal and vertical grade in relation to crash risk is relatively limited too (Wang et al., 2022). The negative parameter of minimum friction (-3.461) indicates that higher minimum friction along the curves decreases the likelihood of crashes (Geedipally et al., 2019). However, the standard deviation of this parameter (1.938) indicates that minimum friction has an

Table 4
Regression results of latent variable random parameters negative binomial (LV-RPNB) model.

	Estimate	SE	t	p value
Constant	−12.653	2.845	−4.448	<0.001
log(AADT)	1.626	0.219	7.437	<0.001
Standard deviation of log(AADT)	0.031	0.033	0.960	0.337
Minimum friction	−4.394	2.086	−2.106	0.035
Standard deviation Minimum friction	1.654	0.313	5.275	<0.001
Super elevation >2%	−1.327	0.508	−2.614	0.009
Positive vertical grade	0.535	0.276	1.935	0.053
Latent variable representing *predictability*	−0.549	0.137	−3.992	<0.001
Structural component:				
Absolute difference in deflection angle	−220.070	4.200	−52.397	<0.001
Curve preceded by straight segment	−30.936	1.458	−21.221	<0.001
Absolute difference in vertical grade	−2.518	1.103	−2.282	0.022
Absolute difference in road width	−2.811	1.262	−2.227	0.026
Sight distance	6.068	4.766	1.273	0.203
σ	0.620	0.181	3.422	0.001
Measurement component:				
α	0.759	0.088	8.657	<0.001
Dispersion parameter	30.000	(fixed for identification)		

increasing effect on crashes on 3.7% of the curves. The dispersion parameter of the negative binomial model (28.365) is also statistically significant indicating that the data are over-dispersed.

Estimation results of the LV-RPNB model (Table 4) show that most of the above independent variables (with almost the same parameters) are statistically significant in this complex model too. However, incorporation of the latent variable representing predictability into this model results in the lack of statistical significance of the parameter for the deflection angle of horizontal curves. The negative parameter of latent variable (−0.549) indicates that higher predictability reduces the likelihood of vehicle crashes along the curves. This finding is consistent with the principles of sustainable safety vision and self-explaining roads (Theeuwes, 2021; Wegman et al., 2008) suggesting that cognitive capabilities of drivers in predicting the road environment contributes to road safety. It is also in line with the past research into design consistency (Dhahir & Hassan, 2019; Lamm et al., 1988) suggesting that abrupt changes in road geometric design contributes to crashes along horizontal curves (David Llopis-Castelló et al., 2018; Mattar-Habib et al., 2008).

The parameter estimates within the structural equation of the latent predictability show that the absolute difference in the deflection angle of horizontal curves has a decreasing effect on predictability. This finding implies that while the deflection angle of horizontal curves may not have a direct statistically significant effect on crashes, larger differences in such deflection angles are associated with lower predictability that, in turn, increases the likelihood of crashes. This finding is in-line with the finding by (Sil et al., 2022) who found that differences in radius and deflection angle between consecutive curves are more likely to influence drivers' ability to distinguish preceding and upcoming curves compared to only the radius and deflection angle of the reference curve. Similarly, the negative parameters of curve preceded by a straight segment, absolute difference in vertical grades, and absolute difference in road width indicate that predictability decreases if there is a significant change in road geometry. Finally, the positive parameter of sight distance indicates that a larger sight distance increases predictability which is in line with the previous findings (Vos et al., 2021a). Research has shown that drivers may perform visual scanning of the environment as a risk-compensating behaviour (Oviedo-Trespalacios et al., 2020), hence larger sight distances increase their predictability and decreases the risk. However, the standard error of this parameter (4.766) indicates that the effect of this variable is not statistically significant for the sample in this study.

Table 5
Comparison of statistical fit between model candidates.

	RPNB	LV-RPNB
N (sample size)	157	157
P (number of parameters)	9	15
LL_0 (null log-likelihood)	−259.484	−538.890
LL (log-likelihood at convergence)	−219.35	−457.74
ρ^2_{adj}(McFadden pseudo-rho squared – adjusted)	0.120	0.123
MAD (mean absolute deviance)	1.390	1.370
MSPE (mean squared predictive error)	11.630	11.130

*Likelihood of the count model component.

During the estimation of the above latent variable model, numerous alternative specifications were investigated too, including a latent variable model specification with the driving behaviour indicators directly and without applying PCA. The results of these models are presented in the appendix and show that none of them provides a statistically significant parameter for the latent variable in the overall Negative Binomial model. This finding implies that each indicator of latent predictability may not be sufficient by and of itself and instead a composite indicator consisting of all indicators obtained via PCA provides a statistically significant parameter for the predictability. It confirms our initial hypothesis that predictability may indeed have several stages and is more complex than can be measured by only one indicator.

5.3. Goodness of statistical fit

Results of global measures of statistical fit for both models are presented in Table 5. The models have very close MAD and MSPE, although the latent variable model has marginally improved measures of statistical fit (about 1% improvement in MAD and about 4% improvement in MSE). These findings indicate that the predictive performance of the two models are comparable. The likelihood-based measure of fit (McFadden pseudo-rho squared) paints a similar picture too ($\rho^2_{adj} = 0.123$ for the latent variable model and $\rho^2_{adj} = 0.120$ for the regular model.

Overall, the findings suggest that incorporating a latent variable into the model does not substantially improve its statistical fit. Nonetheless, the latent variable model provides additional insight about the mechanism of the effect of predictability on crash risk. After all, statistical fit should never be used as the only criterion for model selection. While it may not be obvious to choose one model over the other, both models can be used together to obtain a better understanding of risk factors and their contribution to crashes along horizontal curves.

6. Conclusions

The ability to predict the road plays a crucial role in vehicle crashes along horizontal curves. This predictability can be influenced by roadway geometric design and so it is essential to understand which geometric design elements may cause the road to be (un)predictable. However, predictability may be the result of driver's anticipation and changes in road's geometric design. Driver's anticipation is latent and measuring it may not be trivial. It may also have several stages from the preceding tangent until the midst of the curve. On the top of that, not all drivers anticipate in the same way and thus there may be unobserved heterogeneity in the effect of anticipation on predictability and ultimately on crash risk. Despite methodological advancements in crash risk modelling, there is no econometric model that can adequately explain these complexities. This study aimed to fill this gap by developing an econometric model with a new latent variable that is measured by individual-specific driving behaviour indicators and predicted by location-specific geometric factors.

Results of the proposed econometric model indicated that higher exposure to crashes (in terms of average annual daily traffic) and upward vertical grade were associated with increased likelihood of crashes along horizontal curves. On the contrary, adequate superelevation (more than 2%) decreased the likelihood of these crashes. Similarly, higher pavement friction decreased the likelihood of crashes for most of the curves, but increased this likelihood for a few curves. More importantly, the model showed that the more predictable a curve is, the lower the likelihood of vehicle crashes is along that curve. Although many studies emphasized on this issue in the past, our study provided empirical evidence for the effect of predictability on crash risk. In addition, the results showed that road's predictability is influenced by the differences in the deflection angle of horizontal curves, vertical grades of consecutive road segments, and width of road segments. The findings from the sample in this study showed that larger differences in these factors decrease predictability. As such, providing visual cues for changes in engineering design of the curves may be crucial for the drivers. Finally, the comparison of statistical fit between models with and without latent predictability showed a very small improvement in model performance, implying that the effects of predictability may be accurately estimated using simple observable variables. However, this may also be due to other limitations in our study (further described in the following). Nonetheless, we argue that statistical fit should not be the only criterion to select a model, and that the proposed "model" for predictability corresponding with the hypothesized path diagram presents a closer "picture of reality".

Overall, the results of our analysis can be useful for practitioners and policy makers, as they consolidate the relationship between geometric design features and their contribution to design consistency and eventually predictability. Our findings suggest that roadway engineering design elements are strongly correlated with the unobserved 'predictability' of the road, also taking into account drivers' heterogeneity. At the same time, they indicate the importance of systematically monitoring and accounting for human factors in design policy and practice. In this research, predictability was found to be intuitively related to curve features, but there are a number of human factors that affect the driver interaction with critical road design elements, for which evidence of a clear causal relationships are insufficient, e.g. perception of stopping sight distance in 3D road alignments (Papadimitriou et al, 2018), comprehension of weaving areas (Kusuma et al., 2015), complexity of interchanges (Farah et al., 2017). The latent variable modelling approach in this paper could be used in such questions to provide policy makers with a more complete understanding of human factors in road design.

This study is not without limitations. In measuring latent predictability, we combined the different stages of drivers anticipation in one latent construct and studied the collective effects of those stages on crash risk. Future research should investigate the inter-relationship between those stages prior and during horizontal curves. In addition and for simplicity, we only used the first principal component of driving behaviour indicators for measuring predictability. The marginal improvement in statistical fit in this study may have been due to not considering the other (contributing) components too. Additional driving behaviour data should be collected in future research to present a more complete picture of predictability via all principal components that can add substantially to the explained variance in the data. It should also investigate the effects of using all principal components on models' statistical fit. Some of the explanatory variables (such as sight distance) that were expected to influence road predictability were not statistically significant in the structural equation of the latent variable. Future research should investigate the reasons underlying this lack of statistical significance.

Moreover and due to lack of data availability, we did not evaluate the effects of weather conditions or emerging in-vehicle technologies (e.g. advanced driver assistance systems) on road's predictability and crash risk. Future research should investigate how these factors and these technologies can influence drivers to predict the road and reduce crash risk. Finally, temporal variations were not considered in the effects of explanatory variables on crashes in this study due to lack of proper data.

Table A1
Regression results of latent variable random parameters negative binomial (LV-RPNB) model (without PCA, indicator of predictability = difference between 85th percentile speed at curve start and constant speed point).

	Estimate	SE	t	p value
Constant	−9.842	2.863	−3.438	0.001
log(AADT)	1.445	0.236	6.115	0.000
Standard deviation of log(AADT)	0.021	0.054	0.389	0.697
Minimum friction	−5.350	2.014	−2.657	0.008
Standard deviation Minimum friction	2.100	0.317	6.629	0.000
Super elevation >2%	−1.979	0.508	−3.895	0.000
Positive vertical grade	0.574	0.285	2.015	0.044
Latent variable representing *predictability*	−0.076	0.095	−0.795	0.427
Structural component:				
Absolute difference in deflection angle	−76.324	2.121	−35.991	0.000
Curve preceded by straight segment	−16.590	4.195	−3.954	0.000
Absolute difference in vertical grade	−2.323	1.025	−2.266	0.023
Absolute difference in road width	2.808	4.351	0.645	0.519
Sight distance	142.349	4.327	32.895	0.000
σ	−0.045	0.755	−0.060	0.952
Measurement component:				
α	−0.322	0.080	−4.034	0.000
Dispersion parameter	30.000	(fixed for identification)		

Table A2
Regression results of latent variable random parameters negative binomial (LV-RPNB) model (without PCA, indicator of predictability = Standard deviation of deceleration at maximum deceleration point).

	Estimate	SE	t	p value
Constant	−10.007	2.774	−3.607	0.000
log(AADT)	1.427	0.209	6.838	0.000
Standard deviation of log(AADT)	−0.043	0.030	−1.457	0.145
Minimum friction	−4.825	2.256	−2.138	0.032
Standard deviation Minimum friction	1.950	0.252	7.753	0.000
Super elevation >2%	−1.933	0.465	−4.157	0.000
Positive vertical grade	0.589	0.310	1.901	0.057
Latent variable representing *predictability*	0.101	0.157	0.641	0.522
Structural component:				
Absolute difference in deflection angle	−0.172	2.535	−0.068	0.946
Sight distance	−8.711	6.404	−1.360	0.175
σ	−0.412	1.455	−0.284	0.777
Measurement component:				
α	0.279	0.083	3.364	0.001
Dispersion parameter	30.000	(fixed for identification)		

Table A3
Regression results of latent variable random parameters negative binomial (LV-RPNB) model (without PCA, indicator of predictability = Median distance between start point of deceleration and curve start).

	Estimate	SE	t	p value
Constant	−10.036	2.544	−3.945	0.000
log(AADT)	1.434	0.212	6.756	0.000
Standard deviation of log(AADT)	−0.041	0.026	−1.568	0.117
Minimum friction	−5.007	1.954	−2.562	0.010
Standard deviation Minimum friction	1.950	0.234	8.327	0.000
Super elevation >2%	−1.861	0.491	−3.789	0.000
Positive vertical grade	0.607	0.299	2.028	0.043
Latent variable representing *predictability*	0.077	0.176	0.439	0.661
Structural component:				
Curve preceded by straight segment	3.767	2.460	1.531	0.128
Absolute difference in road width	0.864	0.432	1.998	0.046
σ				
Measurement component:				
α	−0.166	1.426	−0.117	0.907
Dispersion parameter	30.000	(fixed for identification)		

Table A4
Regression results of latent variable random parameters negative binomial (LV-RPNB) model (without PCA, indicator of predictability = Median distance between curve start and constant speed point).

	Estimate	SE	t	p value
Constant	−10.437	1.584	−6.590	0.000
log(AADT)	1.435	0.023	61.663	0.000
Standard deviation of log(AADT)	−0.080	0.018	−4.516	0.000
Minimum friction	−4.356	2.433	−1.791	0.073
Standard deviation Minimum friction	1.477	0.341	4.338	0.000
Super elevation >2%	−1.862	0.559	−3.332	0.001
Positive vertical grade	0.645	0.278	2.318	0.020
Latent variable representing *predictability*	−0.028	0.020	−1.374	0.170
Structural component:				
Absolute difference in deflection angle	2.692	1.580	1.704	0.089
Curve preceded by straight segment	−3.355	2.635	−1.273	0.203
Absolute difference in vertical grade	−15.567	1.778	−8.753	0.000
Absolute difference in road width	1.165	0.027	43.173	0.000
Sight distance	3.150	1.608	1.959	0.050
σ	14.080	7.104	1.982	0.047
Measurement component:				
α	0.016	0.015	1.061	0.289
Dispersion parameter	30.000	(fixed for identification)		

Table A5
Regression results of latent variable random parameters negative binomial (LV-RPNB) model (without PCA, indicator of predictability = Median distance between maximum deceleration point and curve start).

	Estimate	SE	t	p value
Constant	−13.321	1.998	−6.667	0.000
Standard deviation constant	0.931	0.158	5.901	0.000
log(AADT)	1.443	0.196	7.361	0.000
Standard deviation of log(AADT)	−0.071	0.015	−4.838	0.000
Super elevation >2%	−1.439	0.408	−3.524	0.000
Positive vertical grade	0.542	0.281	1.925	0.054
Latent variable representing *predictability*	0.113	0.086	1.317	0.188
Structural component:				
Absolute difference in vertical grade	4.656	4.390	1.061	0.289
Absolute difference in road width	4.915	6.166	0.797	0.425
σ	1.500	1.028	1.460	0.144
Measurement component:				
α	−0.197	0.083	−2.373	0.018
Dispersion parameter	30.000	(fixed for identification)		

However, such temporal instability may exist as a result of external factors such as COVID-19 and the corresponding measures which may alter the effects of the aforementioned explanatory variables.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Appendix A

Table A1–A5.

References

AASHTO. 2001. Policy on geometric design of highways and streets. American Association of State Highway and Transportation Officials, Washington D.C., 1(990), 158.

Afghari, A.P., Washington, S., Haque, M.M., Li, Z., 2018. A comprehensive joint econometric model of motor vehicle crashes arising from multiple sources of risk. Analytic Methods in Accident Research 18, 1–14.

Afghari, A.P., Washington, S., Prato, C., Haque, M.M., 2019. Contrasting case-wise deletion with multiple imputation and latent variable approaches to dealing with missing observations in count regression models. Analytic Methods in Accident Research 24, 100104.

Afghari, A.P., Faghih Imani, A., Papadimitriou, E., van Gelder, P., Hezaveh, A.M., 2021a. Disentangling the effects of unobserved factors on seatbelt use choices in multi-occupant vehicles. Journal of Choice Modelling 41, 100324.

Afghari, A.P., Papadimitriou, E., Li, X., Kaye, S.A., Oviedo-Trespalacios, O., 2021b. How much should a pedestrian be fined for intentionally blocking a fully automated vehicle? A random parameters beta hurdle model with heterogeneity in the variance of the beta distribution. Analytic Methods in Accident Research 32, 100186.

Afghari, A.P., Papadimitriou, E., Pilkington-Cheney, F., Filtness, A., Brijs, T., Brijs, K., Cuenen, A., De Vos, B., Dirix, H., Ross, V., Wets, G., 2022. Investigating the effects of sleepiness in truck drivers on their headway: An instrumental variable model with grouped random parameters and heterogeneity in their means. Analytic Methods in Accident Research 36, 100241.

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention 41 (1), 153–159.

Azmeri Khan, S., Afghari, A.P., Yasmin, S., Haque, M.M., 2023. Applying Random Parameters Negative Binomial Lindley Model to Examine the Effects of Geometric Design Consistency on Run-off Road Crashes Along Two-lane Rural Roadway. Accident Analysis and Prevention 186, 107042.

Bauer, K.M., Harwood, D.W., 2013. Safety effects of horizontal curve and grade combinations on rural two-lane highways. Transportation Research Records 2398 (1), 37–49.

Bhat, C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transportation Research Part B 35 (7), 677–693.

Borsos, A., Birth, S., Vollpracht, H.J. 2015. The role of human factors in road design. 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom).

Brown, T.A., Moore, M.T. 2012. Confirmatory factor analysis. Handbook of structural equation modeling. Rick Hoyle, The Guildford Press, New York, NY.

Campbell, J.L., Richard, C.M., Graham, J., 2008. Human factors guidelines for road systems, No. 600A. NCHRP National Cooperative Highway Research Program, Washington DC.

Chang, L.Y., 2005. Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. Safety Science 43 (8), 541–557.

Charlton, S.G., 2007. The role of attention in horizontal curves: A comparison of advance warning, delineation, and road marking treatments. Accident Analysis and Prevention 39 (5), 873–885.

Charlton, S.G., Starkey, N.J., 2017. Driving on urban roads: How we come to expect the 'correct'speed. Accident Analysis and Prevention 108, 251–260.

Davidse, R.J., Duijvenvoorde, K.V., Louwerse, W.J.R. 2020. Fatal road crashes on national roads in 2019; Analysis of crash and injury factors and resulting potential countermeasures. SWOV, Leidschendam (R-2020-29), 45.

Dhahir, B., Hassan, Y., 2019. Probabilistic, safety-explicit design of horizontal curves on two-lane rural highways based on reliability analysis of naturalistic driving data. Accident Analysis and Prevention 123, 200–210.

Donnell, E., Wood, J., Himes, S., Torbic, D., 2016. Use of side friction in horizontal curve design: A margin of safety assessment. Transportation Research Records 2588 (1), 61–70.

Elvik, R., 2013. International transferability of accident modification functions for horizontal curves. Accident Analysis and Prevention 59, 487–496.

Endsley, M.R., 2017. Toward a theory of situation awareness in dynamic systems. In: Situational awareness. Routledge, pp. 9–42.

Farah, H., van Beinum, A., Daamen, W., 2017. Empirical Speed Behavior on Horizontal Ramp Curves in Interchanges in the Netherlands. Transportation Research Record 2618 (1), 38–47.

Geedipally, S.R., Pratt, M.P., Lord, D., 2019. Effects of geometry and pavement friction on horizontal curve crash frequency. Journal of Transportation Safety and Security 11 (2), 167–188.

Hox, J.J., Bechger, T.M., 1998. An introduction to structural equation modeling. Family Science Review 11 (1998), 354–373.

Huitema, B.E., McKean, J.W., 1991. Autocorrelation estimation and inference with small samples. Psychological Bulletin 110 (2), 291.

Kuiken, M., van der Horst, R., Theeuwes, J., 2012. Designing Safe Road Systems: A Human Factors Perspective. Ashgate Publishing Ltd.

Kusuma, A, Liu, R, Choudhury, C., Montgomery F. 2015. Lane-changing characteristics at weaving section. In: Transportation Research Board 94th Annual Conference, 11-15 Jan 2015, Washington DC. Trasnsportation Research Board of the National Academies.

Lamm, R., Choueiri, E.M., Hayward, J.C., Paluri, A., 1988. Possible design procedure to promote design consistency in highway geometric design on two-lane rural roads. Transportation Research Records 1195, 111.

Lehtonen, E., Lappi, O., Summala, H., 2012. Anticipatory eye movements when approaching a curve on a rural road depend on working memory load. Transportation Research Part F 15 (3), 369–377.

Lewis-Evans, B., Charlton, S.G., 2006. Explicit and implicit processes in behavioural adaptation to road width. Accident Analysis and Prevention 38 (3), 610–617.

Llopis-Castelló, D., Bella, F., Camacho-Torregrosa, F.J., García, A., 2018. New consistency model based on inertial operating speed profiles for road safety evaluation. Journal of Transportation Engineering Part A 144 (4).

Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37 (1), 35–46.
Ma, Q., Yang, H., Wang, Z., Xie, K., Yang, D., 2020. Modeling crash risk of horizontal curves using large-scale auto-extracted roadway geometry data. Accident Analysis and Prevention 144, 105669.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1–16.
Martensen, H., Diependaele, K., Daniels, S., Van den Berghe, W., Papadimitriou, E., Yannis, G., Van Schagen, I., Weijermars, W., Wijnen, W., Filtness, A., Talbot, R., 2019. The European road safety decision support system on risks and measures. Accident Analysis and Prevention 125, 344–351.
Mattar-Habib, C., Polus, A., Farah, H., 2008. Further Evaluation of the Relationship between Enhanced Consistency Model and Safety of TwoLane Rural Roads in Israel and Germany. European Journal of Transport and Infrastructure Research 8 (4).
McLaughlin, S.B., Hankey, J.M., Klauer, C., Dingus, T.A., 2009. Contributing Factors to Run-off-road Crashes and Near-crashes. National Highway Traffic Safety Administration (NHTSA), Washington, DC, USA.
Montella, A., Imbriani, L.L., 2015. Safety performance functions incorporating design consistency variables. Accident Analysis and Prevention 74, 133–144.
Munkin, M.K., Trivedi, P.K., 1999. Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. The Econometrics Journal 2 (1), 29–48.
Nama, S., Sil, G., Maurya, A.K., Maji, A., 2020. Acceleration and deceleration behavior in departing and approaching sections of curve using naturalistic driving data. In: Recent Advances in Traffic Engineering. Springer, pp. 693–704.
Oviedo-Trespalacios, O., Afghari, A.P., Haque, M.M., 2020. A hierarchical Bayesian multivariate ordered model of distracted drivers' decision to initiate risk-compensating behaviour. Analytic Methods in Accident Research 26, 100121.
Papadimitriou, E., Mavromatis, S., Psarianos, B., 2018. Stopping sight distance adequacy assessment on freeways: the case of left horizontal curves over crest vertical curves. Transportation Letters 10 (5), 269–279.
Papadimitriou, E., Filtness, A., Theofilatos, A., Ziakopoulos, A., Quigley, C., Yannis, G., 2019. Review and ranking of crash risk factors related to the road infrastructure. Accident Analysis and Prevention 125, 85–97.
Paschalidis, E., Choudhury, C.F., Hess, S., 2019. Combining driving simulator and physiological sensor data in a latent variable model to incorporate the effect of stress in car-following behaviour. Analytic Methods in Accident Research 22, 100089.
Peng, J., Chu, L., Wang, T., Fwa, T., 2021. Analysis of vehicle skidding potential on horizontal curves. Accident Analysis and Prevention 152, 105960.
Reymond, G., Kemeny, A., Droulez, J., Berthoz, A., 2001. Role of lateral acceleration in curve driving: Driver model and experiments on a real vehicle and a driving simulator. Human Factors 43 (3), 483–495.
Schneider, W.H., Savolainen, P.T., Moore, D.N., 2010. Effects of horizontal curvature on single-vehicle motorcycle crashes along rural two-lane highways. Transportation Research Records 2194 (1), 91–98.
Shaon, M.R., Qin, X., Afghari, A.P., Washington, S., Haque, M.M., 2019. Incorporating behavioral variables into crash count prediction by severity: A multivariate multiple risk source approach. Accident Analysis and Prevention 129, 277–288.
Sil, G., Maji, A., Majumdar, B.B., Maurya, A.K., 2022. Drivers' ability to distinguish consecutive horizontal curves. Canadian Journal of Civil Engineering.
Theeuwes, J., 2021. Self-explaining roads: What does visual cognition tell us about designing safer roads? Cognitive Research 6 (1), 15.
Theeuwes, J., Godthelp, H., 1995. Self-explaining roads. Safety Science 19 (2–3), 217–225.
Tipping, M.E., Bishop, C.M., 1999. Probabilistic principal component analysis. Journal of the Royal Statistical Society 61 (3), 611–622.
Van Petegem, J.W.H., Schermers, G. 2016. Bendiness, European Road Safety Decision Support System, developed by the H2020 project SafetyCube. Retrieved from www.roadsafety-dss.eu on July 2022.
Via, 2020. ViaStat. Retrieved from https://1 via.software. Accessed. 2020.
Vos, E., Bennis, T., Bouman, F., Kuijper, P., Voskuilen, J., Groenendijk, J., 2017. Skid resistance on national roads. Rijkswaterstaat Major Projects and Maintenance (RWS.
Vos, J., Farah, H., Hagenzieker, M., 2021a. How do Dutch drivers perceive horizontal curves on freeway interchanges and which cues influence their speed choice? IATSS research 45 (2), 258–266.
Vos, J., Farah, H., Hagenzieker, M., 2021b. Speed behaviour upon approaching freeway curves. Accident Analysis and Prevention 159, 106276.
Walker, G.H., Stanton, N.A., Chowdhury, I., 2013. Self Explaining Roads and situation awareness. Safety Science 56, 18–28.
Wang, J., He, S., Zhai, X., Wang, Z., Fu, X., 2022. Estimating mountainous freeway crash rate: Application of a spatial model with three-dimensional (3D) alignment parameters. Accident Analysis and Prevention 170, 106634.
Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., 2020. Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC, Boca Raton, FL.
Wegman, F., Aarts, L., Bax, C., 2008. Advancing sustainable safety: National road safety outlook for The Netherlands for 2005–2020. Safety Science 46 (2), 323–343.
Xin, C., Wang, Z., Lee, C., Lin, P.S., Chen, T., Guo, R., Lu, Q., 2019. Development of crash modification factors of horizontal curve design features for single-motorcycle crashes on rural two-lane highways: A matched case-control study. Accident Analysis and Prevention 123, 51–59.
Yasmin, S., Eluru, N., Haque, M.M., 2022. Addressing Endogeneity in modeling Speed Enforcement, Crash Risk and Crash Severity Simultaneously. Analytic Methods in Accident Research 100242.