Accident Analysis and Prevention 179 (2023) 106902

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Short Communication

Further compelling evidence for safety-in-numbers: It is more than meets the eye

Shahram Heydari [a,*], Rune Elvik [b]

[a] Transportation Research Group, Department of Civil, Maritime and Environmental Engineering, University of Southampton, Southampton, UK
[b] Institute of Transport Economics, Gaustadalleen 21, 0349 Oslo, Norway

ARTICLE INFO

Keywords:
Safety-in-numbers
Cross-sectional models
Endogeneity models
Heterogeneity in mean models
Pedestrian
Cyclist

ABSTRACT

In the extant road safety literature, estimating safety-in-numbers is dominated by conventional cross-sectional methods in which active mode (pedestrian or cyclist) volume together with motorised traffic volume are present in regression models explaining active mode safety directly. There is "direct" evidence for safety-in-numbers when the coefficient associated with active mode volume is negative (safety improves as volume increases) or when it is smaller than one (safety decreases at a lower rate compared to the rate of increase in active mode volume). In this article we extend the concept of safety-in-numbers in the traffic safety field, introducing "indirect" safety-in-numbers, which constitutes a new form of evidence for this phenomenon. We provide empirical evidence to support this, discussing that using an approach based on heterogeneity in mean modelling–a form of random parameters (slopes) models–it is possible to reveal "indirect" safety-in-numbers effects. Therefore, such models can reveal further compelling evidence for safety-in-numbers. Accurate knowledge of safety-in-numbers effects (both direct and indirect) and their underlying mechanisms can help provide robust motives for promoting active travel and will have valuable implications for the design of road safety interventions.

1. Introduction

Promoting walking and cycling is considered an important step in achieving healthier population and cleaner environment in urban areas (de Hartog et al., 2010). In this regard, increasing the safety of active modes of travel (walking and cycling) plays a key role (Braun et al., 2016; Khattak and Rodriguez, 2005). Some previous studies indicated that pedestrians and cyclists are safer in locations with higher prevalence of walking and cycling (Jacobsen, 2015), confirming the so-called safety-in-numbers effect. This is perhaps partly because in such locations various road users (e.g., pedestrians, cyclist, and drivers) are well aware of the fact that the road network is shared between different road user types. For example, car drivers expect frequent interactions with pedestrians or cyclists and are thus more cautious, increasing traffic safety. Also, in pedestrian- and cycle-friendly areas, since many drivers or their family members may walk or cycle occasionally, hostile attitudes towards active travel is less pronounced, resulting in an improved safety condition for vulnerable road users (Aldred, 2016).

The concept of safety-in-numbers relates to the idea that the higher the number of pedestrians and/or cyclists, the safer the road network for walking and/or cycling (Elvik and Bjørnskau, 2017; Fyhri et al., 2017; Heydari et al., 2020). Specifically, a safety-in-numbers effect indicates that as the volume of pedestrians or cyclists increases, the safety of these modes of travel decreases at a lower rate compared to that of their respective exposures (Elvik and Bjørnskau, 2017). For example, a value of 0.8, which is smaller than one, for the coefficient associated with pedestrian volume (in a model that explains pedestrian crash frequency) indicates a safety-in-numbers effect although this estimated coefficient implies that pedestrian volume and the number of crashes are positively associated. A negative association, however, indicates a strong safety-in-numbers effect, meaning that the safety of active modes increases (e.g., crash frequency decreases) as pedestrian and/or cyclist activity (volume) increases (Elvik and Goel, 2019). Literature provides evidence for both safety-in-numbers and strong safety-in-numbers effects at both micro (e.g., intersection) and macro (area) levels (Aldred et al., 2019; Aldred et al., 2018; Elvik, 2016; Heydari et al., 2017; Lee et al., 2019; Murphy et al., 2017; Tasic et al., 2017; Xu et al., 2019). While reviewing the safety-in-numbers literature is beyond the scope of this paper, for a comprehensive systematic review, see Elvik and Goel (2019).

* Corresponding author.
E-mail addresses: s.heydari@soton.ac.uk (S. Heydari), Rune.Elvik@toi.no (R. Elvik).

https://doi.org/10.1016/j.aap.2022.106902
Received 13 July 2022; Received in revised form 25 October 2022; Accepted 15 November 2022
Available online 21 November 2022

1.1. The current paper

In this article we discuss a new form of safety-in-numbers that can be inferred indirectly from certain statistical models and explain how such an indirect safety-in-numbers effect can be revealed. We briefly revisit issues of endogeneity in the context of safety-in-numbers since this serves to identify the strength of the evidence for safety-in-numbers. We compare different statistical models with respect to the strength and form of evidence (direct and/or indirect) they provide for safety-in-numbers, introducing the safety-in-numbers triangle. We conclude with a summary and provide some recommendations for future research in this specific area, with the overarching aim of promoting walking and cycling.

2. Strength of evidence for safety-in-numbers: moderate versus strong evidence

The strength of evidence for safety-in-numbers is under question as previous studies mostly analysed cross-sectional data using traditional statistical models (e.g., negative binomial regression), revealing the magnitude of the association between active modes traffic volume and their safety (Elvik and Bjørnskau, 2017). In fact, such traditional models are unable to account for endogeneity of explanatory variables (Elvik, 2021; Mannering et al., 2020). This may be exacerbated by unobserved heterogeneity (Heydari, 2018; Mannering et al., 2016), which if not addressed properly, would result in bias in estimation.

Ideally, stronger evidence for safety-in-numbers can be shown using approaches that can address endogeneity in the effect of pedestrian or cyclist flow on the safety of pedestrians or cyclists. However, the use of endogeneity models in the context of safety-in-numbers is rare, if not non-existent, in the extant road safety literature. We believe endogeneity is particularly important in this context where the focus is on interpreting the estimated coefficients associated with pedestrian and/or cyclist volumes. Therefore, the reliability of such interpretations matters particularly. Some claim that safety-in-numbers really is numbers-in-safety, i.e., people walk or cycle more once a safe infrastructure is provided. It is therefore not unlikely that pedestrian or cycle volume is endogenous. So, if the analysis does not accommodate endogeneity, statistical inferences would be biased, the extent of which is unclear and depends on the application and data.

Selectivity bias can arise easily here; for example, less safety-conscious drivers (e.g., those who often engage in risk-taking behaviours such as speeding) may avoid locations (intersections, road segments, neighbourhoods, etc.) with higher prevalence of walking and cycling to avoid lower operating speeds. Therefore, the crash data collected at such locations may not be a truly random and representative sample of the driving population. Consequently, an estimated parameter for traffic exposure (pedestrian/cyclist and traffic volumes) may be biased if selectivity is not accommodated.

3. Further compelling evidence for safety-in-numbers: direct versus indirect evidence

Previous research mostly employed traditional statistical models that only allow the analyst to examine the “direct” association between pedestrian or cyclist volume and their respective safety conditions. That is, active modes volume appears in count regression models as an explanatory variable explaining pedestrian or cyclist crash frequencies directly. For example, consider the generic random parameters (slopes) count regression model specified in Eq. (1). The regression coefficient β, associated with active mode volume, being smaller than one indicates the presence of a safety-in-numbers effect.

$$y_j \sim Poisson\left(\lambda_j\right)$$

$$log\left(\lambda_j\right) = \alpha_0 + \alpha \ln(AADT_j) + \beta \ln(activemodevolume_j) + \boldsymbol{\gamma X_j} + \eta_j Z_j + \in_j$$

$$Z_j \sim normal\left(\mu_z, v_z\right) \tag{1}$$

where y is the observed number of crashes involving pedestrian or cyclists for sites j; λ is the expected crash frequency and is a function of traffic volume (AADT), active mode (pedestrian or cyclist) volume, and some other site characteristics $\boldsymbol{X}$ and Z. α is the regression coefficient associated with motorised traffic volume; $\boldsymbol{X}$ represents the vector of other contributory factors with their respective parameters γ; and α_0 is an intercept. For simplicity, we assume only the effect of one contributory factor, Z, varies across the sample with its associated random parameter η; μ_z and v_z are respectively the mean and variance of the random parameter Z. And $\in$ is a random term, which captures overdispersion in crash datasets. Under both conventional cross-sectional and endogeneity approaches, the model reveals only “direct” evidence for safety-in-numbers as the analyst can directly investigate the association between active mode volume and safety.

3.1. Indirect evidence via heterogeneity in mean models

Recently, heterogeneity in mean models have been employed in the crash literature when using random parameters (slopes) models. Besides addressing unobserved heterogeneity more fully, these models can explain the source of heterogeneity in the data better than conventional random parameters models (Heydari et al., 2018; Seraneeprakarn et al., 2017). Note that the heterogeneity in means approach is known to add robustness to random parameters models, leading to more reliable statistical inferences. Specifically, heterogeneity in mean models can explain the variability in the effect of random regression parameters (i.e., within covariate variability) based on explanatory variables $\boldsymbol{W}$, which are available in the data; i.e., $Z_j = f(\boldsymbol{W_j})$. The vector of covariates $\boldsymbol{W}$ can include any set of explanatory variables that are available in the data. For simplicity, let's assume $\boldsymbol{W}$ includes only active mode volume; a generic heterogeneity in mean specification can be written as in Eq. (2).

$$y_j \sim Poisson\left(\lambda_j\right)$$

$$log\left(\lambda_j\right) = \alpha_0 + \alpha \ln(AADT_j) + \beta ln(activemodevolume_j) + \boldsymbol{\gamma X_j} + \eta_j Z_j + \in_j$$

$$Z_j \sim normal\left(\mu_{zj}, v_z\right)$$

$$\mu_{zj} = a_0 + a_1(active\ mode\ volume_j) \tag{2}$$

In the above formulation, let's assume Z is a risk factor, having an undesirable impact on safety. Since its mean depends on active mode volume, as specified in Eq. (2), an “indirect” safety-in-numbers effect can be revealed by identifying the indirect association through Z between active mode volume and its safety. A negative sign for a_1 implies that an increase in active mode volume decreases the adverse effect of this risk factor on safety; and consequently, it indicates a “strong” indirect safety-in-numbers effect. Also, a_1 being positive with a value smaller than one (i.e., when the adverse effect of a risk factor increases less than in proportion to pedestrian or cyclist volume) indicates an indirect safety-in-numbers effect. Similarly, in case Z, instead of being a risk factor, has a protective effect on safety, an indirect safety-in-numbers effect is revealed when pedestrian or cyclist counts have a positive sign in the function explaining Z.

If the coefficient associated with active mode volume (β in Eq. (1)) is a random parameter, its mean will indicate whether a conventional direct safety-in-numbers effect exists overall. The mean being smaller than zero indicates a strong direct safety-in-numbers effect and the mean being positive but smaller than one indicates a direct safety in numbers effect. Its variance will then indicate the uncertainty around this effect. A study conducted by (Xu et al., 2021) discusses how the variance of a random parameter affects the accuracy of out-of-sample crash predictions; however, in the context of safety-in-numbers usually prediction is not of interest. Alternatively, if a heterogeneity in mean specification is considered for this parameter (β in Eq. (2)), the presence (and magnitude) of a safety-in-numbers effect dependents on other explanatory variable(s) affecting the variation in β. In the latter scenario, the heterogeneity in mean specification for β can reveal a direct safety-in-numbers effect as the analyst can observe the direct effect of active mode volume on pedestrian/cyclist safety.

3.2. Empirical evidence for an "indirect" safety-in-numbers effect

A study conducted by Heydari et al., 2020 is the first and only example that reveals and discusses "indirect" evidence for safety-in-numbers based on a heterogeneity in mean model. They investigated the varying effect of proximity to school on pedestrian injury frequencies at signalised intersections in Montreal. They found that the adverse effect of proximity to school, a major risk factor, decreased as pedestrian volume increased at intersections in proximity to schools (intersection located in a 400 m buffer around schools). In other words, a higher prevalence of pedestrians at intersections around schools resulted in a lower rate of increase in pedestrian injury frequencies at these intersections. Considering the above notation, in Heydari et al., 2020 pedestrian volume was among $\boldsymbol{W}$, capturing the variability in the effect of proximity to school on pedestrian safety. This indicated a strong "indirect" safety-in-numbers effect since the sign of the coefficient associated with pedestrian volume, explaining the effect of proximity to school, was found to be negative, −0.152 (see Table 4 in Heydari et al., 2020). With respect to the direct (conventional) safety-in-numbers effect, the same study found direct evidence for safety-in-numbers as well. This is implied since the coefficient associated with pedestrian volume (being an explanatory variable among others in explaining pedestrian injury frequency as in Eq. (1)) was 0.417. That is, a value smaller than one, indicating that safety decreases less than in proportion to pedestrian volume.

To better clarify the above discussion, Fig. 1 displays a schematic view of both direct and indirect evidence for safety-in-numbers according to Heydari et al., 2020. The discussion above and the evidence provided by the authors indicate that safety-in-numbers is not only what has been revealed in previous research (direct evidence), but it can also manifest in other forms, providing indirect evidence for safety-in-numbers. That is, higher levels of pedestrian and cyclist activity may improve safety indirectly while affecting other factors that have a bearing on the safety of active modes; for example, by hampering their adverse effects.

3.3. Heterogeneity in variance specification

Note that a further extension of the heterogeneity in mean specification allows the variance of the random parameter Z to be modelled as a function of other explanatory variables in the data, indicating that uncertainty around the effect of Z varies systematically across the data. Pedestrian or cyclist volume having a bearing on the variance of Z means they govern the variance of the effect of Z on pedestrian or cyclist safety. A positive sign will indicate the effect of Z becomes more dispersed (uncertain) as pedestrian or cyclist counts increase and vice versa. While in this case, active mode volume provides insights into the precision of the effect of Z on safety, it does not imply a safety-in-numbers effect although it might be considered favourable when precision is increased. Therefore, we do not discuss heterogeneity in variance models further.

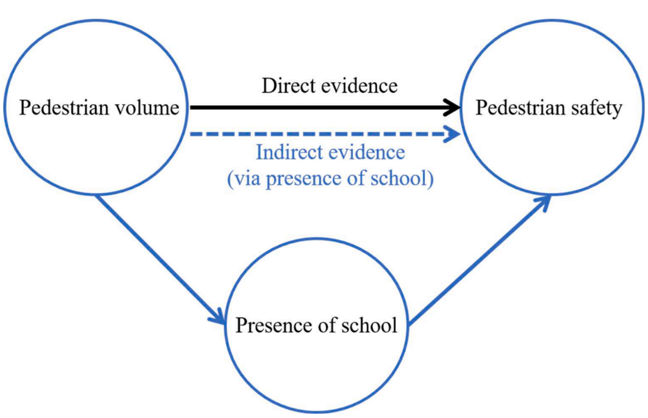

Fig. 1. Example of direct and indirect evidence for safety-in-numbers. Note that pedestrian volume does not affect the presence of school, but it affects its effect on pedestrian safety.

3.4. Models with interaction terms

Note that indirect safety-in-numbers effects can be inferred from models that include interaction terms as well. For example, consider the model defined in Eq. (3), where there is an interaction term between active mode volume and the explanatory variable X. Also, let's assume X is a continuous variable.

$$y_j \sim Poisson\left(\lambda_j\right)$$

$$log\left(\lambda_j\right) = \alpha_0 + \alpha \log(AADT_j) + \beta log(active\,mode\,volume_j) + \gamma_1 X_j + \gamma_2 X_j log(active\,mode\,volume_j) + \epsilon_j \tag{3}$$

where γ_1 and γ_2 are regression coefficients associated with X and the interaction variable. In the above specification, the effect of X on log*(λ)*, for any fixed value of log*(active mode volume)* such as log*(active mode volume₀)*, can be obtained from (4)

$$\gamma_1 + \gamma_2 log(activemodevolume_0) \tag{4}$$

Therefore, the effect of X depends on active mode volume. This implies that if γ_2 is smaller than one, the effect of X on log*(λ)* increases less than in proportion to active mode volume, indicating an "indirect" safety-in-numbers effect. This is indirect since active mode volume comes into play indirectly by affecting the impact of another variable on safety. However, we do not discuss models with interaction terms further as their use is very limited in the crash literature. This is perhaps partly due to the fact that correlation between the explanatory variables, in particular, with interaction terms can be very high, causing serious collinearity issues.

4. Safety-in-numbers triangle

To better clarify how different statistical models differ from each other in terms of form and strength of evidence for safety-in-numbers, we introduce the "safety-in-numbers triangle" in Fig. 2. Given the focus of our study, we group statistical models into four main categories: traditional cross-sectional models, traditional endogeneity models (see Mannering et al. (2020) for further details), heterogeneity in means models, and heterogeneity in means endogeneity models. Here, in the context of safety-in-numbers, both traditional and heterogeneity in mean endogeneity models refer to those models that account for the endogeneity of pedestrian and cyclist volume. Note, however, that any form of model specification that improves the reliability of statistical inferences contributes favourably to the strength of evidence for safety-in-numbers. This said, a heterogeneity in means endogeneity models in its most ideal form in this context not only can reveal both direct and indirect safety-in-numbers effects, but it can also address the potential endogeneity of active mode volume in Eq. (2). This would allow for a more reliable estimation of β (see Eq. (2)) which in turn will increase the strength of evidence for a safety-in-numbers effect.

Fig. 2 displays different types of statistical models in terms of their capability in revealing differing forms of evidence (direct vs indirect) for safety-in-numbers and the strength of evidence they are able to provide.

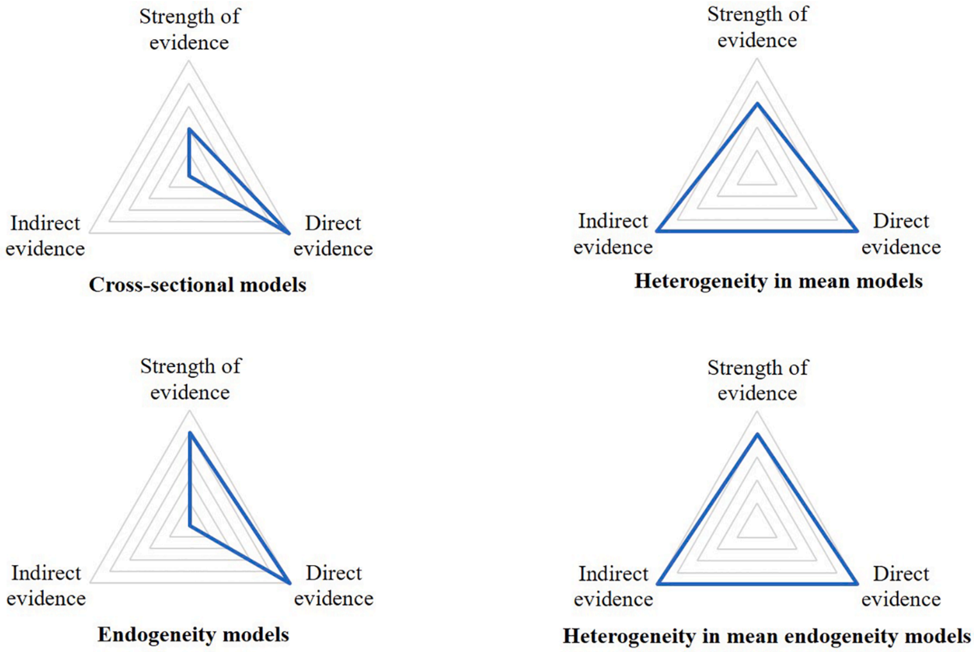

Fig. 2. Safety-in-numbers triangle, comparing different models with respect to the strength and form of evidence. Note that here endogeneity models primarily is intended to refer to those that address endogeneity of pedestrian or cyclist volume.

Both categories of traditional cross-sectional and endogeneity models can reveal only direct evidence while their heterogeneity in means counterparts can reveal both direct and indirect evidence. Therefore, the latter approaches are superior to traditional models, being able to draw a fuller picture. Besides requiring more detailed data and more in-depth investigation, heterogeneity in means and endogeneity models are often more cumbersome computationally and more time consuming compared to traditional cross-sectional models. Therefore, the analyst must trade-off between model complexity, revealing indirect/direct effects, and the strength of evidence for safety-in-numbers. As can be seen in Fig. 2, endogeneity models provide stronger evidence for safety-in-numbers effects compared to traditional cross-sectional models. Based on the four categories of models considered in Fig. 2, heterogeneity in mean endogeneity models are the most comprehensive ones.

5. Summary and recommendations

In general, a safety-in-numbers effect is observed when the safety of active modes of travel decreases at a lower rate compared to the rate of increase in active modes volume. A strong safety-in-numbers effect is observed when the safety of active modes increases as active modes exposure increases. The safety-in-numbers effect in its traditional definition investigates the “direct” impact of pedestrian or cyclist volumes on pedestrian or cyclist safety, respectively. However, this definition does not fully appreciate the potential safety benefits of higher levels of walking and cycling. In this paper, we introduced and discussed a new form of safety-in-numbers based on “indirect” evidence, while discussing other important concerns such as endogeneity in this context.

Conventional safety-in-numbers studies often employ cross-sectional statistical models with the aim of revealing “direct” safety-in-numbers effects. In such studies pedestrian or cyclist volumes are present in regression models as explanatory variables. Conclusions are then drawn regarding the presence of safety-in-numbers effects by estimating the association between these exposure terms and safety. However, an ideal study of safety-in-numbers would aim at revealing both direct and indirect evidence for safety-in-numbers while addressing possible selection-bias and endogeneity in all explanatory variables, especially, those relating to active modes exposure variables. While providing further compelling evidence, revealing both direct and indirect evidence for safety-in-numbers provides a fuller picture, assisting the analyst in extracting more meaningful information that can in turn lend itself to safety and urban planning policy. In this regard, further research is thus needed in the traffic safety field.

In addition to the capability of heterogeneity in mean models in better addressing unobserved heterogeneity (and therefore, improving statistical inference), their applications should be further investigated in the context of safety-in-numbers. Also, if the focus of a study is on revealing safety-in-numbers, we strongly recommend employing endogeneity models to avoid bias induced by selectivity and endogeneity. Obviously, this necessitates the data being collected in a way that allows for accommodating endogeneity, an important factor to consider at an early stage of a study. This said, depending on research question(s) to be answered in a study, the analyst can decide on trade-offs associated with various statistical models regarding model complexity, data availability, and the strength and form (direct vs indirect) of evidence for safety-in-numbers effects. To conclude, safety-in-numbers, whether in its direct or indirect form as we discussed in this paper, can have important implications for road safety policy, especially where strong evidence seems to exist. This is because by promoting and increasing walking and cycling activities, authorities can achieve safer road networks for vulnerable road users particularly in urban areas, provided other adequate infrastructure and countermeasures are in place.

CRediT authorship contribution statement

Shahram Heydari: Conceptualization, Investigation, Methodology, Visualization, Writing – review & editing, Writing – original draft. **Rune Elvik:** Conceptualization, Investigation, Methodology, Writing – review & editing, Writing – original draft.

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

No data was used for the research described in the article.

References

Aldred, R., 2016. Cycling near misses: Their frequency, impact, and prevention. Transp. Res. A Policy Pract. 90, 69–83. https://doi.org/10.1016/j.tra.2016.04.016.

Aldred, R., Goodman, A., Gulliver, J., Woodcock, J., 2018. Cycling injury risk in London: A case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits. Accid. Anal. Prev. 117, 75–84. https://doi.org/10.1016/j.aap.2018.03.003.

Aldred, R., Goel, R., Woodcock, J., Goodman, A., 2019. Contextualising Safety in Numbers: a longitudinal investigation into change in cycling safety in Britain, 1991–2001 and 2001–2011. Inj. Prev. 25 (3), 236. https://doi.org/10.1136/injuryprev-2017-042498.

Braun, L.M., Rodriguez, D.A., Cole-Hunter, T., Ambros, A., Donaire-Gonzalez, D., Jerrett, M., Mendez, M.A., Nieuwenhuijsen, M.J., de Nazelle, A., 2016. Short-term planning and policy interventions to promote cycling in urban centers: Findings from a commute mode choice analysis in Barcelona, Spain. Transp. Res. A Policy Pract. 89, 164–183. https://doi.org/10.1016/j.tra.2016.05.007.

de Hartog, J.J., Boogaard, H., Nijland, H., Hoek, G., 2010. Do the health benefits of cycling outweigh the risks? Environ. Health Perspect. 118 (8), 1109–1116.

Elvik, R., 2016. Safety-in-numbers: Estimates based on a sample of pedestrian crossings in Norway. Accid. Anal. Prev. 91, 175–182. https://doi.org/10.1016/j.aap.2016.03.005.

Elvik, R., 2021. Why are there so few experimental road safety evaluation studies: Could their findings explain it? Accid. Anal. Prev. 163, 106467.

Elvik, R., Bjørnskau, T., 2017. Safety-in-numbers: A systematic review and meta-analysis of evidence. Saf. Sci. 92, 274–282. https://doi.org/10.1016/j.ssci.2015.07.017.

Elvik, R., Goel, R., 2019. Safety-in-numbers: An updated meta-analysis of estimates. Accid. Anal. Prev. 129, 136–147. https://doi.org/10.1016/j.aap.2019.05.019.

Fyhri, A., Sundfør, H.B., Bjørnskau, T., Laureshyn, A., 2017. Safety in numbers for cyclists—conclusions from a multidisciplinary study of seasonal change in interplay and conflicts. Accid. Anal. Prev. 105, 124–133. https://doi.org/10.1016/j.aap.2016.04.039.

Heydari, S., 2018. A flexible discrete density random parameters model for count data: Embracing unobserved heterogeneity in highway safety analysis. Anal. Methods Accid. Res. 20, 68–80. https://doi.org/10.1016/j.amar.2018.10.001.

Heydari, S., Fu, L., Miranda-Moreno, L.F., Jopseph, L., 2017. Using a flexible multivariate latent class approach to model correlated outcomes: A joint analysis of pedestrian and cyclist injuries. Anal. Methods Accid. Res. 13, 16–27. https://doi.org/10.1016/j.amar.2016.12.002.

Heydari, S., Fu, L., Thakali, L., Joseph, L., 2018. Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: Applications to grade crossing safety analysis. Anal. Methods Accid. Res. 19, 33–48. https://doi.org/10.1016/j.amar.2018.06.003.

Heydari, S., Miranda-Moreno, L., Hickford, A.J., 2020. On the causal effect of proximity to school on pedestrian safety at signalized intersections: A heterogeneous endogenous econometric model. Anal. Methods Accid. Res. 26, 100115 https://doi.org/10.1016/j.amar.2020.100115.

Jacobsen, P.L., 2015. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj. Prev. 21 (4), 271. https://doi.org/10.1136/ip.9.3.205rep.

Khattak, A.J., Rodriguez, D., 2005. Travel behavior in neo-traditional neighborhood developments: A case study in USA. Transp. Res. A Policy Pract. 39 (6), 481–500. https://doi.org/10.1016/j.tra.2005.02.009.

Lee, J., Abdel-Aty, M., Xu, P., Gong, Y., 2019. Is the safety-in-numbers effect still observed in areas with low pedestrian activities? A case study of a suburban area in the United States. Accid. Anal. Prev. 125, 116–123. https://doi.org/10.1016/j.aap.2019.01.037.

Mannering, F., Bhat, C.R., Shankar, V., Abdel-Aty, M., 2020. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. Anal. Methods Accid. Res. 25, 100113 https://doi.org/10.1016/j.amar.2020.100113.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16. https://doi.org/10.1016/j.amar.2016.04.001.

Murphy, B., Levinson, D.M., Owen, A., 2017. Evaluating the Safety In Numbers effect for pedestrians at urban intersections. Accid. Anal. Prev. 106, 181–190. https://doi.org/10.1016/j.aap.2017.06.004.

Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., Milton, J., 2017. Occupant injury severities in hybrid-vehicle involved crashes: A random parameters approach with heterogeneity in means and variances. Anal. Methods Accid. Res. 15, 41–55. https://doi.org/10.1016/j.amar.2017.05.003.

Tasic, I., Elvik, R., Brewer, S., 2017. Exploring the safety in numbers effect for vulnerable road users on a macroscopic scale. Accid. Anal. Prev. 109, 36–46. https://doi.org/10.1016/j.aap.2017.07.029.

Xu, P., Xie, S., Dong, N., Wong, S.C., Huang, H., 2019. Rethinking safety in numbers: are intersections with more crossing pedestrians really safer? Inj. Prev. 25 (1), 20. https://doi.org/10.1136/injuryprev-2017-042469.

Xu, P., Zhou, H., Wong, S., 2021. On random-parameter count models for out-of-sample crash prediction: Accounting for the variances of random-parameter distributions. Accid. Anal. Prev. 159, 106237 https://doi.org/10.1016/j.aap.2021.106237.