Accident Analysis and Prevention 179 (2023) 106899

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Driver mileage and accident involvement: A synthesis of evidence

Rune Elvik

Institute of Transport Economics, Gaustadalleen 21, 0349 Oslo, Norway

ARTICLE INFO

Keywords:
Drivers
Mileage
Accident involvement
Non-linearity
Meta-analysis
Evidence synthesis

ABSTRACT

The relationship between driver mileage and accident involvement has been a controversial topic for at least 20 years. The key issue is whether driver accident involvement rate increases in proportion to miles driven or has a non-linear relationship to miles driven. This paper presents a synthesis of evidence from studies of how the number of accidents per driver per unit of time relates to distance driven in the same period. Most studies of this relationship are methodologically weak and their results highly inconsistent and potentially misleading. Unreliable data and poor control for confounding factors characterise most studies. Only a few studies based on multivariate statistical models control for at least some of the confounding factors that may influence the relationship between distance driven and accident involvement. These studies consistently show that the number of accidents per driver per year increases less than in proportion to distance driven. A good approximation is that the number of accidents per driver per unit of time is proportional to the square root of distance driven. Potential methodological and substantive explanations of this finding are discussed.

1. Introduction

The use of accident rates to control for the effect on the number of accidents of distance driven has a long history in road safety research. An accident rate (accidents divided by distance driven) implies a linear relationship between distance driven and the number of accidents: if you drive twice as long, you can expect twice as many accidents. However, since about 1995 (Hauer, 1995), there has been increasing doubt about the validity of accident rates. With respect to driver accident rates, studies showing a non-linear relationship between mileage and the number of accidents started to appear around 1990 (Janke, 1991). A paper in 2002 by Hakamies-Blomqvist et al. (2002) set off a discussion about so called "low mileage bias". Low mileage bias refers to the biased impression that older drivers have higher accident rates than other drivers, when in fact their accident rates are no higher than other drivers with a low annual mileage. The discussion about low-mileage bias has continued and no consensus has emerged (Staplin et al., 2008; Langford et al., 2008; af Wåhlberg, 2009).

The objective of this paper is to systematically review and summarise the findings of studies of the relationship between drivers' annual driving distance and accident involvement rate (accidents per driver per unit of time). Three main questions are asked: (1) Is the relationship linear or non-linear? (2) If non-linear, what is the shape of the relationship? (3) What could be the explanations of a non-linear relationship?

2. Literature survey

Sciencedirect and Google Scholar were searched using "mileage" AND "accidents" OR "crashes" as search terms. There were no language restrictions and no limits on study age. 16 studies containing relevant data were identified. Table 1 lists these studies in chronological order.

For each study, Table 1 lists the year it was published; the country where it was conducted; the groups of drivers included; the source of accident data; the source of exposure data; how mileage was operationally defined (as a variable stating driving distance in intervals or as a continuous variable); the number of distance categories used if distance was stated by means of intervals; how data were analysed and the number of confounding variables controlled for in analysis.

The studies were published between 1991 and 2021. Eight countries are represented, all of them high-income and highly motorised. All studies included non-professional drivers of both genders. Some studies included drivers of all ages, other studies included only young or old drivers. In most studies, both accidents and mileage were self-reported by drivers. Two main approaches to analysis of data are found in the studies. The first is to tabulate data on driving distance and accident rate and fit curves to the data. The second is to develop multivariate statistical models to explain variation in driver accident involvement.

Table 2 gives an example of tabulated data. The table is taken from Forsyth et al. (1995). It shows data for females during their three first years of driving. Mileage is stated as an interval. There are seven

E-mail address: re@toi.no.

https://doi.org/10.1016/j.aap.2022.106899
Received 30 August 2022; Received in revised form 23 October 2022; Accepted 9 November 2022
Available online 14 November 2022

intervals. The driving distance at the midpoint of each interval is stated at the bottom of the Table. There are two estimators of driver accident involvement: (1) Accidents per driver per year, (2) Accidents per million miles driven.

The two estimators of accident involvement move in opposite directions. As driving distance increases, the number of accidents per driver per year also increases. The number of accidents per million miles driven, however, goes down. For reasons explained in the next section of the paper, the best estimator of driver accident involvement is the number of accidents per driver per unit of time.

To summarise tabulated data like those shown in Table 2, one may fit curves to the data points. Fig. 1 shows an example of this. A curve was fitted to the data on accidents per driver per year for females in their second year of driving, shown in Table 2.

A power function best fitted the data. It closely tracks the data points. The goodness-of-fit statistics (R-squared) is therefore close to its maximum value of 1.

3. Methodological aspects of the studies

There are two types of studies of the relationship between driving distance and driver accident involvement. One type of study tabulates accident rates for drivers belonging to different intervals for annual driving distance. Table 2 and Fig. 1 illustrated this type of study. The other type of study develops multivariate statistical models to explain variation in driver accident involvement. Most of the studies listed in Table 1 are of the first type. In particular, the tradition of "low-mileage-bias" studies started by Hakamies-Blomqvist et al. (2002), tabulated data on accidents per million vehicle kilometres of driving for drivers belonging to three intervals for driving distance: (1) <3000 km, (2) between 3000 and 14,000 km, and (3) more than 14,000 km (Langford et al., 2006; 2008; 2013; Alvarez and Fierro, 2008).

All "low-mileage bias" studies found that accident rate declined sharply as annual driving distance increased. There are, however, several problems with these studies. These problems make their results highly uncertain and possibly misleading. The most important problems are:

1. Driving distance, and in some studies, accidents are self-reported by drivers. Self-reported data have been found to be inaccurate.
2. Each study has few data points, between 3 and 7. This limits the possibilities of testing different functional forms describing the relationship between driving distance and accident involvement.
3. The studies do not control for any confounding variables, except for age. In most studies, fairly large age groups (e.g. 31–64) are used.
4. The definition of exposure (driving distance) and accident involvement (accidents per million kilometres of driving) can generate a spurious negative relationship between the variables.

Staplin et al. (2008) found that low-mileage drivers understated their

Table 1
List of studies.

Study	Country	Groups of drivers included	Source of accident data	Source of exposure data	Estimate of driving distance	Number of distance categories	Approach to data analysis	Confounders controlled for
Janke, 1991	United States	All ages	State records	Self-reported	Midpoints of intervals	7	Curves fitted to tabulated data	0
Maycock and Lockwood, 1993	Great Britain	All ages	Self-reported	Self-reported	Midpoints of intervals and exact values	6 or continuous variable	Curves fitted to tabulated data and multivariate statistical model	8 in multivariate analysis
Forsyth et al., 1995	Great Britain	Newly licensed (most aged < 30)	Self-reported	Self-reported	Midpoints of intervals and exact values	7 or continuous variable	Curves fitted to tabulated data and multivariate statistical model	7 in multivariate analysis
Assum, 1997	Norway	All ages	Self-reported	Self-reported	Midpoints of intervals	7	Curves fitted to tabulated data	1
Massie et al., 1997	United States	All ages	Official statistics	Self-reported	Exact values	Continuous variable	Multivariate statistical model	8 in multivariate analysis
Lourens et al., 1999	Netherlands	All ages	Self-reported	Self-reported	Midpoints of intervals	5	Curves fitted to tabulated data	1
Hakamies-Blomqvist et al., 2002	Finland	Ages 26–40 and 65-	Self-reported	Self-reported	Mean values for intervals	3	Curves fitted to tabulated data	1
Fontaine, 2003	France	All ages	Official statistics	Self-reported	Midpoints of intervals	3	Curves fitted to tabulated data	1
Langford et al., 2006	Netherlands	All ages	Not stated	Self-reported	Mean values for intervals	3	Curves fitted to tabulated data	1
Alvarez and Fierro, 2008	Spain	All ages	Self-reported	Self-reported	Mean values for intervals	3	Curves fitted to tabulated data	1
Ferreira and Minikel, 2012	United States	All ages	Insurance claims	Odometer data	Exact values	Continuous variable	Multivariate statistical model	2
Boucher et al., 2013	Spain	Young drivers (<40)	Insurance claims	Odometer data	Exact values	Continuous variable	Multivariate statistical models	1 to 3
Langford et al., 2013	Australia	Older drivers (70-)	Self-reported	Self-reported	Midpoints of intervals	3	Curves fitted to tabulated data	1
Antin et al., 2017	United States	Older drivers (65-)	Naturalistic driving data	Naturalistic driving data	Mean values for intervals; exact values	7 or continuous variable	Curves fitted to tabulated data and multivariate statistical model	2 in multivariate model
Hua et al., 2018	Australia	Older drivers (75-)	Self-reported	Naturalistic driving data	Mean values for intervals	3	Curves fitted to tabulated data	1
Elvik, 2021	Norway	Age 24	Self-reported	Self-reported	Exact values	Continuous variable	Multivariate statistical model	4

driving distance. High-mileage drivers tended to overstate their driving distance. The true difference in driving distance between high-mileage and low-mileage drivers is therefore smaller than the self-reported difference. Langford et al. (2008) argued that the errors were not large enough to eliminate the tendency for accident rate to decline with increasing annual driving distance. It was, however, large enough to create bias in any curve or coefficient fitted to describe this relationship. Af Wåhlberg and Dorn (2015) show that the test–retest reliability of self-reported data on mileage, violations and accidents is low. In other words, drivers do not, for example, always report the same driving distance when asked about it at intervals of 3 to 7 months. Therefore, data recorded by insurance companies or public authorities are preferred to self-reported data. In the studies relying on insurance data, the focus is on accidents, which means that other events that may be reported to insurance companies, such as theft or vandalism, are unlikely to be included in the data.

Most of the "low-mileage bias" studies have only three data points (three intervals for driving distance). This means that a second-degree polynomial will fit the data points perfectly and always be preferred if different functional forms are assessed according to their goodness of fit. Most common functional forms – linear, power, logarithmic, exponential or polynomial – will fit well when there are few data points. The parameters of these functions will be spuriously precise.

Several factors influence the relationship between driving distance and accident involvement. In particular, low-mileage drivers and high-mileage drivers differ with respect to the traffic environments they do most of their driving in. Table 3 shows some the differences that have been found in some studies.

Forsyth et al. (1995) found that male low-mileage drivers do 9 % of their driving on motorways. High-mileage drivers do 24 % of their driving on motorways. Since motorways have a low accident rate, part of the reason why high-mileage drivers have a lower accident rate than low-mileage drivers could be that they do more of their driving on the safest roads.

Keall and Frith (2006) found that low-mileage drivers do more of their driving on urban roads in daytime than high-mileage drivers. The net effect of this on accident rate is not clear. On the one hand, accident rate tends to be lower in daytime than at night. On the other hand, urban roads have a higher accident rate than rural roads, at least with respect to less serious accidents. Hanson and Hildebrand (2011), like Forsyth et al. (1995), found that high-mileage drivers do more of their driving on highways (which presumably includes motorways) than low-mileage drivers.

These differences between low-mileage and high-mileage drivers mean that it is important to control for type of traffic environment when estimating the effect of driving distance on accident involvement. Only studies based on multivariate statistical models can do this well. In studies based on cross-tabulation, drivers can be divided into groups which are homogeneous with respect to confounding variables, but this rapidly reduces sample size.

Finally, the definition of risk and exposure in the "low-mileage bias" studies can generate a spurious negative relationship between the variables. Risk is estimated as the number of accidents per kilometre driven (A/B). Exposure is defined as the number of kilometres driven by each driver (B/C). In other words, risk equals A/B and exposure equals B/C.

It is obvious that defining exposure and risk this way can generate a spurious negative relationship between exposure and risk. Consider what happens when B increases. All else equal, the value of A/B will decrease, i.e. risk is reduced. When B increases, the value of B/C increases, i.e. exposure increases at the same time as risk decreases. There is thus, by definition, a negative relationship between exposure and risk. Elvik (2013) generated random numbers for traffic volume and accidents for 159 junctions and showed that a negative relationship between exposure and risk could arise from these random numbers.

There does not have to be a negative relationship between driving distance and accident rate; the result stated above applies only if A (the number of accidents) and C (the number of drivers) are kept constant while B (kilometres per driver) increases. However, to avoid ambiguity, it is better to estimate accident involvement as the number of accidents per driver per unit of time. This definition of accident involvement is consistent with the use of count regression models, like negative binomial regression, in developing statistical models to explain variation in the number of accidents between units of observation (drivers, road sections, bridges, etc.). Furthermore, it may be noted that the count of accidents usually has a known statistical distribution (Poisson, negative binomial, Poisson-lognormal, etc.) in a sample of drivers, whereas accident rates (accidents per mile) have no known distribution.

In sum, the problems associated with the "low-mileage bias" studies support rejecting these studies as potentially erroneous because of unreliable data, few data points, poor control of confounding variables and an unsuitable definition of the variable measuring accident involvement. The synthesis of evidence therefore only includes studies employing multivariate count regression models.

4. Data extraction from studies included in synthesis

The studies that are candidates for a formal synthesis of their findings by means of meta-analysis are Maycock and Lockwood (1993), Forsyth et al. (1995), Massie et al. (1997), Ferreira and Minikel (2012), Boucher et al. (2013), Antin et al. (2017) and Elvik (2021). All these studies have the following characteristics:

1. A multivariate statistical model was developed to explain variation in accident involvement, controlling for at least some of the confounding factors that may influence the relationship between driving distance and accident involvement.
2. The mathematical formulations of the models are similar or allow for similar interpretations of the regression coefficients.
3. Driving distance was included as a continuous variable, which can take on any positive value.

Whether these similarities are sufficient to make the estimated regression coefficients for driving distance comparable is discussed in the next section. From each of the studies, one or more regression coefficients referring to driving distance, and the standard error of the coefficient, were extracted.

Table 2
Example of tabulated data. Taken from Forsyth et al., 1995.

Year of driving	Females	Annual mileage < 1000	1000–2999	3000–4999	5000–6999	7000–9999	10000–14999	15,000 or more
First	Accidents per year	0.101	0.145	0.162	0.195	0.197	0.279	0.260
	Accidents per million miles	202.0	72.5	40.5	32.5	24.3	22.3	13.0
Second	Accidents per year	0.055	0.093	0.119	0.128	0.139	0.175	0.227
	Accidents per million miles	110.0	46.5	29.8	21.3	17.4	14.0	11.4
Third	Accidents per year	0.055	0.082	0.091	0.099	0.147	0.172	0.175
	Accidents per million miles	110.0	41.0	22.8	16.5	18.4	13.8	8.8
	Annual mileage (midpoints)	500	2000	4000	6000	8000	12,500	20,000

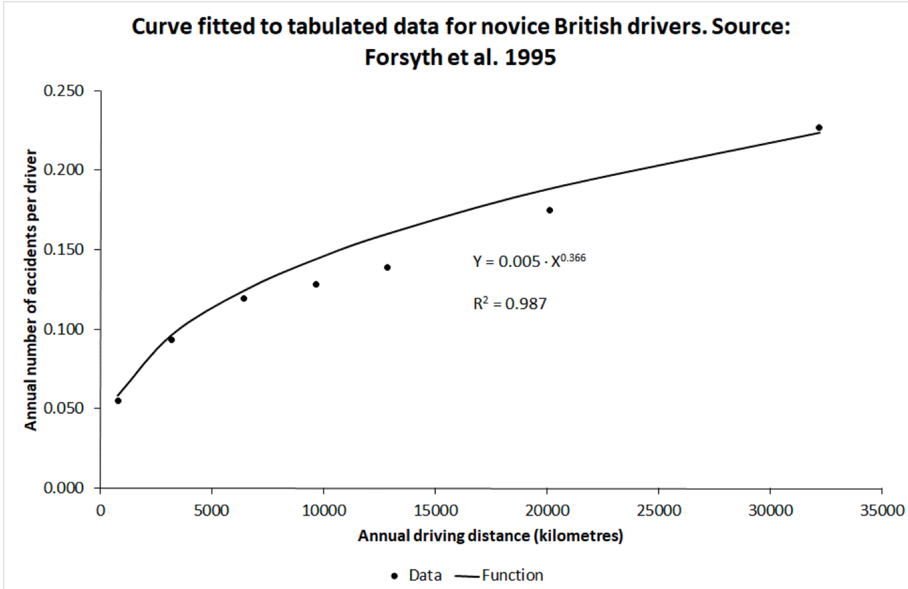

Fig. 1. Example of curves fitted to tabulated data. Female drivers in their second year of driving.

Table 3
Differences between low-mileage and high-mileage drivers with respect to traffic environment.

Study	Group	Percent of driving (by time or distance) done in different traffic environments: Driving environment	Lowest mileage group	Highest mileage group
Forsyth et al., 1995	Males	Motorways	9	24
		Quiet parts of towns or cities	29	15
	Females	Motorways	6	22
		Quiet parts of towns or cities	30	13
Keall and Frith, 2006	15–19 year olds	Urban roads in daytime	49	36
	25–64 year olds	Urban roads in daytime	46	29
	65 or older	Urban roads in daytime	62	24
Hanson and Hildebrand, 2011	All	Major or urban highways	10	18

Maycock and Lockwood (1993) fitted a model with the following form, using accidents per driver per year as the dependent variable:

$$\text{Accidents per driver per year} = \beta_0 \bullet Mileage^{\beta_1} \bullet e^{\left(\sum_{i=1}^{n}\beta_n \bullet X_n\right)} \quad (1)$$

This model formulation is common and, with minor differences, used by all multivariate studies reviewed below. Mileage is raised to a power, estimated by β_1. This coefficient can be interpreted as an elasticity, i.e. it shows the percentage change in the number of accidents per driver per unit of time when driving distance increases by one percent. The effects of the confounding variables are captured by the exponential function (e raised to the expression in parenthesis). The strength of the model developed by Maycock and Lockwood is that it controls for many confounding variables: age, gender, driving experience, socio-economic group, driving in darkness, driving on motorways and driving on rural roads. The weakness is that both accidents and mileage were self-reported. One regression coefficient (0.279) and its standard error (0.016) were extracted from the study.

The model developed by Forsyth et al. (1995) had the same mathematical form as that of Maycock and Lockwood (1993). It controlled for age, gender, driving experience, trip frequency and driving in rain. Data on accidents and mileage were self-reported. One regression coefficient (0.572) and its standard error (0.039) were extracted from the study.

Massie et al. (1997) developed models predicting accident rates (accidents per million miles driven). Three models were developed: one for fatal accidents, one for injury accidents and one for property-damage accidents. The models were fitted to groups of drivers formed according to age (63 groups), gender (2 groups) and time of driving (day vs night; 2 groups). Thus, each model was fitted to a data set consisting of 63 • 2 • 2 = 252 observations. One observation was lost because there were no travel data for females aged 85 or older. Travel data were taken from the national household travel survey. Accident data were taken from official statistics. The coefficients for mileage in these models show risk elasticities. They were converted to accident elasticities by adding 1. Standard errors were assumed not to be affected by this transformation. The models controlled for age, gender and time of driving and included interaction terms between these variables. Three regression coefficients (0.6585; 0.5049; 0.5398) and their standard errors (0.0336; 0.0354; 0.0324) were extracted from the study.

Ferreira and Minikel (2012) developed models based on pay-as-you-drive insurance data for the state of Massachusetts. The data set included more than 2.8 million car insurance years. All data were taken from insurance records. The final model included mileage and dummies for territory and insurance class. Insurance class was based on driver characteristics and type of use of the car. Unfortunately, the standard errors of regression coefficients are not reported, but it is stated that in a model including only mileage, the coefficient was significant at the 2 • 10^{-16} level. This suggests a very low standard error. As an approximation, it has been assumed to equal the standard error of the number of claims used in the analysis, divided by the number of claims. This gives an estimate of 0.0025 for the standard error. The data on mileage in this study are presumably correct, as they come from pay-as-you-drive insurance. Minor damage may not be reported to the insurance company, but otherwise the claims data should also be accurate. One regression coefficient (0.40) and the estimate of its standard error (0.0025) were extracted from the study. Although the accuracy of the standard error is unknown, it is considerably smaller than in the other studies reviewed in this section, consistent with having a low p-value.

Boucher et al. (2013) also relied on pay-as-you-drive insurance data in their analysis. The data are limited to drivers below the age of 40. The models developed controlled for age, gender, car age, and if the car is parked in a garage. Four regression coefficients (0.4188; 0.4903; 0.5505; 0.4908) and their standard errors (0.0418; 0.0389; 0.0967; 0.0860) were extracted from the study.

Antin et al. (2017) fitted models to data for participants in the SHRP-2 naturalistic driving study. The data both on mileage and accidents are presumably accurate and complete in this study, as cars participating in the naturalistic driving study were fitted with instruments recording driving distance and accidents. The models used accident rate as dependent variable. The coefficient for mileage in the model for total accidents has been converted to an accident elasticity by adding 1. The study controlled for age group and gender. One regression coefficient (0.48) and its standard error (0.12) were extracted from the study.

Finally, Elvik (2021) fitted a model to a sample of 24-year old drivers. Mileage and accidents were self-reported. The model controlled for age, gender, county, and how urbanised the place of residence was. One regression coefficient (0.193) and its standard error (0.028) were extracted from the study.

5. Exploratory meta-analysis

To determine whether a formal synthesis of the regression coefficients extracted from the studies makes sense, four issues must be resolved:

1. Determining the comparability of the regression coefficients
2. Assessing the potential presence of publication bias
3. Assessing the degree of heterogeneity in coefficient estimates
4. Assessing the potential presence of outlying data points.

5.1. Comparability of regression coefficients

Card (2012) argues that a formal synthesis of regression coefficients only makes sense if the regression models are identical in all respects, i.e. have the same mathematical form and include identical sets of variables, identically measured. His main reason for imposing this restriction is that estimates of regression coefficients are often sensitive to the variables included. Hence, two regression coefficients may differ, not because the underlying effect differs, but because different sets of variables were included in the models.

Hauer (2010) proposes a solution to the problem identified by Card (2012). If several regression models, including different variables, have been fitted to a data set, the estimates of the coefficient for a specific variable can be compared across the different model specifications. If the estimates do not differ much, they are comparable. Stability of the value of a regression coefficient across different model specifications suggests that the coefficient is robust with respect to confounding, i.e. its estimated value is not much influenced by which potentially confounding variables a model includes. Elvik and Bjørnskau (2017) show examples of how regression coefficients can be compared across different model specifications. It should be noted that comparing different models developed in a single study is analogous to comparing different studies, each of which developed only a single model. An advantage of comparing different models developed in the same study, is that these models are based on the same data. One may therefore rule out the possibility that variation in the estimates of a regression coefficient for a certain variable is attributable to differences in the data used to estimate the coefficient.

Unfortunately, most of the studies discussed above include just a single version of the regression model. Comparing the regression coefficient of principal interest, the coefficient for mileage, across models is therefore not possible for all studies. However, some studies permit a comparison between curves fitted to tabulated data and regression coefficients estimated in multivariate models. The curves fitted to tabulated data control for fewer confounding variables than the multivariate statistical models.

To illustrate this approach, consider the study by Maycock and Lockwood (1993). In addition to the multivariate statistical model, the study presents tabulated data of accidents per driver per year for drivers belonging to different intervals for driving distance. Based on these data, power functions like the one shown in Fig. 1 can be estimated. The power coefficients can then be compared to the estimate of power in the multivariate model. Fig. 2 shows four estimates of power estimated from tabulated data in the study of Maycock and Lockwood, the weighted mean value of these estimates and the coefficient estimated in the multivariate model.

Fig. 2 shows 95 % confidence intervals surrounding each estimate of power. It is seen that the confidence intervals for the estimates based on tabulated data are much wider than the confidence interval based on the multivariate model. Each of the estimates based on tabulated data control for one confounding variable only: gender. The multivariate model controls for several confounding variables.

To evaluate the comparability of the estimates of power, the following comparisons have been made:

Difference in estimates = Estimate based on tabulated data – estimate based on multivariate model.

$$\text{Standard error of difference of estimates} = \sqrt{SE_i^2 + SE_m^2}$$

Thus, the first coefficient listed in Fig. 2 had a value of 0.256. The coefficient estimated in the multivariate model had a value of 0.279. The difference in value is 0.256 – 0.279 = –0.023. The standard errors of the coefficients were, respectively, 0.056 and 0.016. Hence, the standard error of the difference between the coefficients is 0.058 ($\sqrt{0.056^2 + 0.016^2}$). The difference divided by its standard error is.

–0.023/0.058 = –0.395. The conventionally applied critical value for regarding the difference as statistically significant is ± 1.96. The observed value is less than this. There is no statistically significant difference between the coefficients. They are therefore regarded as comparable. Repeating the procedure for the other coefficients listed in Fig. 2 leads to the conclusion that they can all be treated as comparable.

Six coefficients, each controlling for gender and experience, were fitted to tabulated data based on the study by Forsyth et al. (1995). The differences between each of these coefficients and the coefficient estimated in the multivariate model were studied, applying the procedure explained above. Four of the six coefficients were found to be statistically significantly different from the coefficient of the multivariate model. However, the comparison is problematic, as the coefficients do not refer to identically defined variables. The coefficients of the curves fitted to tabulated data refer to mileage. The coefficient estimated in the multivariate model refers to mileage plus driving frequency (number of trips per unit of time).

The four models developed by Boucher et al. (2013) differ slightly from one another. Based on the four models, a weighted mean coefficient was estimated and compared to the other four coefficients. No statistically significant differences were found.

Antin et al. (2017) developed one model for all accidents and one model for at-fault accidents. There was no statistically significant difference between the coefficients estimated in these two models.

Finally, Elvik (2021) developed the model in four stages, at each stage adding another variable. The coefficients for mileage estimated in the three first stages were compared to the coefficient estimated in the final model. No statistically significant differences were found.

In total, 18 comparisons of coefficients were made. In 14 cases, there were no statistically significant differences between the coefficients. In the four cases where a statistically significant difference was found, the coefficients did not refer to identically defined variables. It is concluded that the regression coefficients are sufficiently comparable for a formal synthesis of them to make sense.

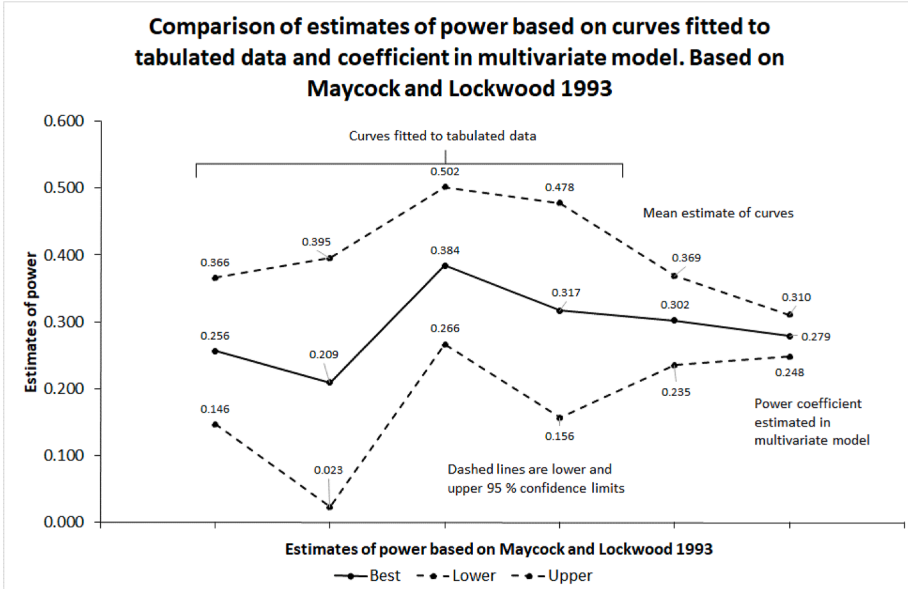

Fig. 2. Comparison of estimates of power based on curves fitted to tabulated data and coefficient in multivariate model.

5.2. Testing for publication bias

Fig. 3 shows a funnel plot of estimates of regression coefficients. The axes have been defined as recommended by Sterne and Egger (2001). The horizontal axis shows the natural logarithm of the estimate of a regression coefficient. The vertical axis shows the fixed-effects standard error of each coefficient. Each data point is an estimate of a regression coefficient for mileage.

If there is no publication bias, the distribution of the data points should resemble a funnel turned upside down, with the narrow end (small dispersion of data points) at the top and the wide end (large dispersion of data points) at the bottom. The data points in Fig. 3 do not quite have such a distribution. Moreover, the weighted mean estimate of the regression coefficients is strongly influenced by the uppermost data point in the diagram. This is the estimate from Ferreira and Minikel (2012), who did not state the standard error. A rough estimate was given above, consistent with their explanation that the coefficient had a very low p-value, which implies a low standard error. However, the accuracy of the estimate is unknown.

Testing for publication bias by means of the trim-and-fill method (Duval and Tweedie, 2000a, 2000b; Duval, 2005) indicates a slight publication bias. Three data points were trimmed away. This had a negligible effect on the weighted mean estimate of the regression coefficients, whose value changed only by 0.4 %. Stronger publication bias is indicated if the study by Ferreira and Minikel (2012) is omitted. However, the results are highly implausible as the value of trimmed weighted mean regression coefficient is the second lowest of the individual estimates. Again, a single study (Maycock and Lockwood, 1993) strongly influences results, because it has a much higher statistical weight than any of the other studies. The test that included all studies is regarded as best and it indicated that any publication bias had a negligible influence.

The problem of a single study having a much larger statistical weight than any other study is reduced by adopting a random-effects model of

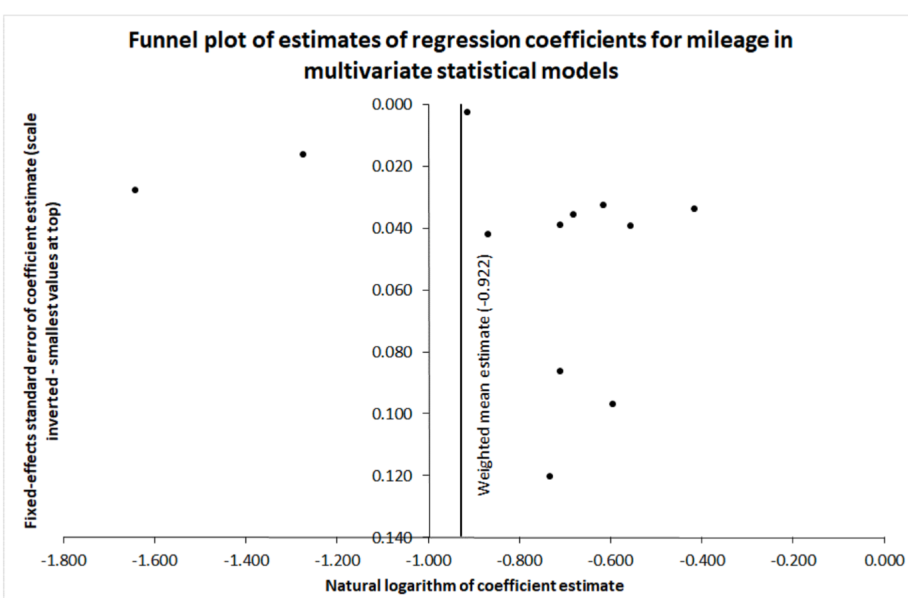

Fig. 3. Funnel plot of estimates regression coefficient for mileage in multivariate statistical models.

analysis. This is the appropriate model when the estimates to be synthesised are widely dispersed. The coefficient estimates for mileage are very heterogeneous. If estimates are weighted according to a random-effects statistical weight (see details in next section), no estimate contributes more than 10 % to the sum of the statistical weights. A trim-and-fill analysis relying on random-effects statistical weights indicates a slight publication bias. Three data points are trimmed away, and the value of the trimmed weighted mean estimate of the regression coefficient for mileage is reduced by 9.5 %.

5.3. Heterogeneity of coefficient estimates

A formal synthesis of a set of estimates of an effect only makes sense if the estimates are not too heterogeneous. Unfortunately, there is no precise criterion or rule for determining when there is too much heterogeneity in a data set (Elvik, 2018).

In the inverse-variance technique for meta-analysis, each estimate of effect (each regression coefficient for mileage) is assigned a statistical weight, which is inversely proportional to its variance:

$$\textit{Fixed} - \textit{effect statistical weight} \;=\boldsymbol{W}=\frac{1}{\boldsymbol{SE}^2} \tag{2}$$

SE is the standard error of each regression coefficient. This statistical weight accounts for random sampling variation in estimates only, i.e. it assumes that there is only random variation between estimates. To determine whether this is correct, the following statistics are computed:

$$Q=\sum_{i=1}^{g} W_i \bullet Y_i^2-\frac{\left(\sum_{i=1}^{g} W_i \bullet Y_i\right)^2}{\sum_{i=1}^{g} W_i} \tag{3}$$

In formula (3) W is the fixed-effects weight assigned to each estimate, and Y is the estimate. The Q statistic measures variance. It is used to estimate a variance component (τ^2) which is an estimator of the systematic between-study variation in estimates. The variance component is estimated as follows:

$$\text{Variance component }(\tau 2)\;=\frac{Q-(g-1)}{C} \tag{4}$$

Q is defined above, g is the number of estimates. C is estimated as:

$$C\;=\sum_{g=1}^{n} w_i-\left(\frac{\sum_{g=1}^{n} w_i^2}{\sum_{g=1}^{n} w_i}\right) \tag{5}$$

The random-effects statistical weight becomes:

$$\text{Random effects statistical weight}\;=\frac{1}{SE_i^2+\tau^2} \tag{6}$$

For the 12 regression coefficients included in this study, Q was estimated as 229.01 and τ^2 as 0.0112. The I^2 statistic (Borenstein et al., 2009), which indicates the share of systematic variation in the data set, is 95.2 %.

There is, in other words, great heterogeneity. The regression coefficients range in value from 0.193 to 0.659. This range is not particularly large. The coefficients all have the same sign, and all indicate that the number of accidents per driver increases less than in proportion to mileage. Perhaps the most comparable data set is the regression coefficients summarised by Elvik and Goel (2019) in a study of safety-in-numbers. In that study, the regression coefficients for motor vehicle volume (which are comparable to the coefficients in the present study) ranged in value from 2.19 to −1.16. The coefficients for cycle volume ranged from 0.87 to −0.14. The coefficients for pedestrian volume ranged from 1.40 to 0.07. All these ranges are larger than the range of estimates in this study. It is concluded that it makes sense to synthesise the estimates of regression coefficients for mileage based on multivariate statistical models.

5.4. Outlying data points

To test for outlying data points, the summary estimate of the regression coefficient for mileage (random-effects model) was re-estimated by omitting one estimate at a time. If the summary estimate based on N – 1 was outside the 95 % confidence interval for the estimate based on N, the omitted estimate was classified as outlying. No outlying estimates were found.

6. Synthesis of multivariate studies

Table 4 presents five summary estimates of the coefficient for mileage, as estimated in multivariate statistical models. The two first estimates are the simple mean and median; these are not based on meta-analysis. The other three estimates are based on meta-analysis.

All estimates are close to one another and close to the value of 0.5. A coefficient of 0.5 means that the number of accidents per driver per unit of time increases in proportion to the square root of distance driven.

Fig. 4 shows the relationship between the coefficient for mileage and the number of confounding variables controlled for in statistical analysis. There is a weak positive relationship. This means that the value of the coefficient for mileage increases slightly as studies control for more confounding variables.

Unfortunately, the studies that control for many confounding variables rely on self-reported data about mileage and accidents. The studies that have high-quality data on mileage and accidents are Ferreira and Minikel (2012), Boucher et al. (2013) and Antin et al. (2017). The summary coefficient for mileage based on these studies is identical (0.492) to the summary coefficient based on studies that control for at least 8 confounding variables.

7. Potential explanations of the relationship

It is clear that driver accident involvement does not increase in proportion to distance driven. What can explain the tendency for each additional kilometre driven to become safer as more kilometres are driven?

In principle, there are two types of explanation: methodological and substantive. A methodological explanation would be that poor data, e.g. unreliable data on mileage and accidents, or confounding variables, e.g. long-distance drivers doing a larger share of their driving on safe roads, explain the relationship. However, as noted above, regression coefficients for mileage with values less than one have been found in all studies reviewed above. It does not seem to matter much whether the studies rely on self-reported data or on more objective data. Nor does it make much difference how well the studies control for confounding variables. It is clearly still possible that all studies could be wrong. A study based both on high-quality data and controlling for many confounding variables does not exist. Hence, rejecting all studies is an

Table 4
Summary estimates of regression coefficient for mileage in multivariate statistical models.

Summary estimate of power coefficient	Best estimate	Standard error
Simple mean value of 12 coefficients estimated in multivariate models	0.465	0.037
Median value of 12 coefficients estimated in multivariate models	0.491	0.037
Random-effects mean value of 12 coefficients estimated in multivariate models	0.457	0.034
Random-effects mean value of 4 coefficients estimated in multivariate models controlling for 8 confounding variables	0.492	0.055
Random-effects mean value of 6 coefficients estimated in multivariate models relying on high-quality data	0.492	0.051

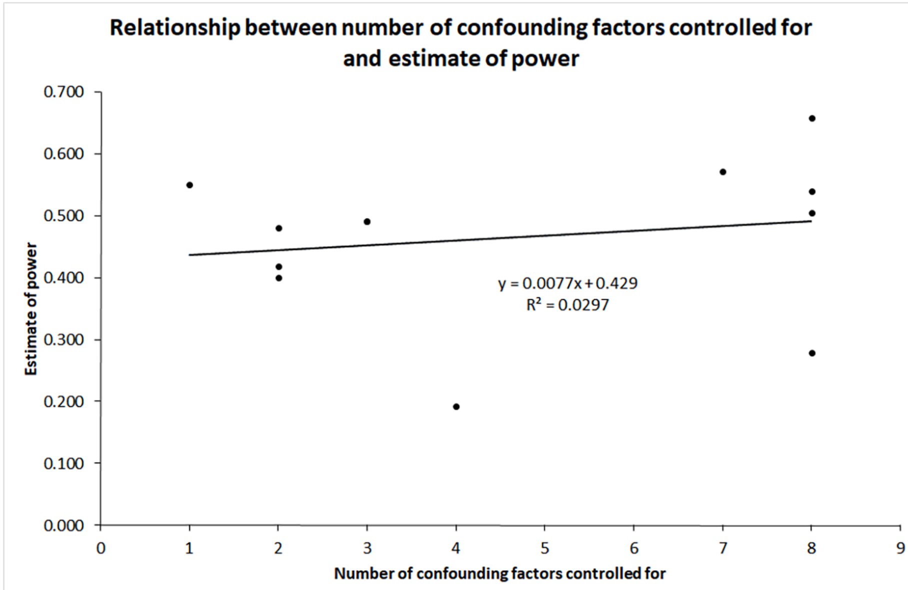

Fig. 4. Relationship between number of confounding variables controlled for and estimate of power.

interpretation that can be defended.

Nevertheless, there are studies suggesting substantive explanations. Driving can be regarded as a life-long process of learning, in which those who drive long distances get more occasions for learning than those who drive short distances. Elvik (2015) considers some implications of defining exposure as events, and notes that the number of events of a specific type, like encounters or simultaneous arrivals in junctions, tend to increase faster than traffic volume. Applied to driving distance, this means that long-distance drivers encounter more rare events than short distance drivers.

Mitroff and Biggs (2014) and Biggs, Adamo and Mitroff (2014) note that events occurring very rarely are noticed less often than events occurring more frequently. They analysed big data from the Airport scanner game. In this game, players earn points by detecting forbidden items in luggage. Some of the items are easy to detect, others may be partly hidden beneath legal items. The forbidden items occurred with a frequency between 0.078 % and 4.14 %, i.e. they were all quite rare. It was found that items occurring with a frequency of<0.5 % were rarely detected. Items occurring with a frequency of 1 % or more were almost always detected.

In traffic, events occur with varying frequency. However, the probability that a driver will experience a rare event is not proportional to his or her driving distance. To see this, image than on a given trip, an event has a probability of 0.00078 of occurring (equal to the lowest probability in the airport scanner game). On a given trip, the event either occurs or it does not. Each trip can be modelled as a binomial trial with two outcomes: event or no event. Suppose a low-distance driver makes 300 trip per year. The probability that he or she will experience the rare event during these 300 trips (300 binomial trials) is shown in Table 5.

Table 5
Increase in probability of experiencing a rare event as a function of the number of trips.

Probability of a rare event per trip = 0.00078	**Number of trips taken per year**				
Probability of N (=0,1,2, etc) rare events per year	**300**	**600**	**900**	**1200**	**1500**
0	0.791	0.626	0.495	0.392	0.310
1	0.185	0.293	0.348	0.367	0.363
2	0.022	0.069	0.122	0.172	0.213
3	0.002	0.011	0.029	0.054	0.083
4	0.000	0.001	0.005	0.013	0.024
5		0.000	0.001	0.002	0.006
6			0.000	0.000	0.001
7					0.000
Sum of probabilities	1.000	1.000	1.000	1.000	1.000
Probability of at least two rare events	0.023	0.081	0.156	0.241	0.326
Probability of at least three rare events	0.002	0.012	0.034	0.069	0.114
Relative distance per year	1.000	2.000	3.000	4.000	5.000
Relative probability of at least two rare events	1.000	3.444	6.686	10.284	13.953
Relative probability of at least three rare events	1.000	6.779	19.343	38.775	64.245

The probability that a driver making 300 trips per year will never experience the rare event is 0.791. The probability that he or she will experience the rare event at least twice is 0.023. Now consider a driver making 1500 trips per year. The probability that this driver will experience the rare event at least twice is 0.326. While making only 5 times as many trips as the low-mileage (few trips) driver, the high-mileage (many trips) driver can expect to encounter the rare event at least twice almost 14 times as often as the low-mileage driver (see bottom of Table 5). This makes the rare event much more predictable and less surprising to the high-mileage driver than to the low-mileage drivers. One may reasonably assume that the high-mileage driver develops better skills, and thereby a lower risk of accident, in dealing with the rare event than the low-mileage driver.

Conversely, low-mileage drivers may lose skills by not practising them often enough to retain them. Many of the skills involved in driving are of a "use-it-or-lose-it" quality. Skills that have become automated, may have to be re-learnt following a period of no or little driving. Naturalistic driving studies could shed light on this, e.g. by showing whether some types of mistakes or poorly executed tasks are made more often by low-mileage drivers than by high-mileage drivers.

8. Discussion

The most important factor influencing a driver's accident involvement is his or her exposure to traffic risk, i.e. how often, how long and where a driver drives. Yet, obvious as this is, there are surprisingly few, and surprisingly poor, studies of how the number of accidents per driver per unit of time is related to the distance driven. Indeed, the literature survey made for this paper identified fewer studies than was found in a similar review of studies of safety-in-numbers (Elvik and Goel, 2019). Although some studies may have been missed, the survey at least includes all the best-known and most cited studies from the past 30 years.

Distance driven is of course not the only factor influencing driver accident involvement. Accident involvement is related to age, gender, experience, and the type of traffic environment where driving takes place. To identify the specific contribution of distance, a study should control for the effects of all the other factors. Few studies do so.

All the studies belonging to the "low-mileage-bias" tradition were rejected. These studies all rely on self-reported data of unknown accuracy and do not employ multivariate methods for analysis. There are few studies relying on high-quality data about driving distance and accident involvement. However, such data sets, which can be quite large, are increasingly available as a result of pay-as-you-drive insurance schemes. Although traditional insurance remains dominant, it is not too bold to predict that pay-as-you-drive insurance will become more widespread. This will improve the quality of the data used to study the relationship between driving distance and accident involvement.

The study reported in this paper should be repeated a few years from now, when better data are likely to be available on a wider scale.

9. Conclusions

The main results of the research presented in this paper can be summarised as follows:

1. There are few studies of the relationship between a driver's annual driving distance and his or her accident involvement.
2. Most studies rely on self-reported data of unknown accuracy and fail to control for confounding factors.
3. Studies employing multivariate statistical models, and relying on high-quality data, suggest that the number of accidents per driver per unit of time increases roughly in proportion to the square root of distance driven.
4. The probability that a driver will encounter a rare event during driving increases much faster than the number of kilometres driven.

CRediT authorship contribution statement

Rune Elvik: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing.

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

Data will be made available on request.

Acknowledgement

This study did not receive funding from any private or public organisation.

References

Alvarez, F.J., Fierro, I., 2008. Older drivers, medical condition, medical impairment and crash risk. Accid. Anal. Prev. 40, 55–60.
Antin, J.J., Guo, F., Fang, Y., Dingus, T.A., Perez, M.A., Hankey, J.M., 2017. A validation of the low mileage bias using naturalistic driving study data. J. Saf. Res. 63, 115–120.
Assum, T., 1997. Attitudes and road accident risk. Accid. Anal. Prev. 29, 153–159.
Biggs, A.T., Adamo, S.H., Mitroff, S.R., 2014. Rare, but obviously there: Effects of target frequency and salience on visual search accuracy. Acta Psychol. 152, 158–165.
Borenstein, M., Hedges, L.V., Higgins, J.T., Rothstein, H., 2009. Introduction to Meta-analysis. John Wiley and Sons, New York.
Boucher, J.-P., Pérez-Marin, A.M., Santolino, M., 2013. Pay-as-you-drive insurance: the effect of the kilometres on the risk of accident. Anales del Instituto de Actuarios Espanoles 19, 135–154.
Card, N.A., 2012. Applied meta-analysis for the social science research. The Guilford Press, New York.
Duval, S., 2005. The trim and fill method. In: Rothstein, H., Sutton, A.J., Borenstein, M. (Eds.), Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments, 127–144. John Wiley and Sons, Chichester.
S. Duval R. Tweedie Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analysis Journal of the American Statistical Association 95 2000a 89 98.
S. Duval R. Tweedie A non-parametric trim and fill method of assessing publication bias in meta-analysis Biometrics 56 2000b 455 463.
Elvik, R., 2013. Can a safety-in-numbers and a hazard-in-numbers effect co-exist in the same data? Accid. Anal. Prev. 60, 57–63.
Elvik, R., 2015. Some implications of an event-based definition of exposure to the risk of road accident. Accid. Anal. Prev. 76, 15–24.
Elvik, R., 2018. Meta-analytic methods. Chapter 19. In: Lord, D., Washington, S. (Eds.), Safe Mobility: Challenges, Methodology and Solutions. Emerald Publishing, Bingley UK, pp. 425–447.
Elvik, R., Bjørnskau, T., 2017. Safety-in-numbers: A systematic review and meta-analysis of evidence. Saf. Sci. 92, 274–282.
Elvik, R., Goel, R., 2019. Safety-in-numbers: An updated meta-analysis of evidence. Accid. Anal. Prev. 129, 136–147.
Elvik, R. (2021). The predictive accuracy of a model of young car driver accidents. Working paper 51802. Oslo, Institute of Transport Economics.
Ferreira, J., Minikel, E., 2012. Measuring per mile risk for pay-as-you-drive automobile insurance. Transp. Res. Rec. 2297, 97–103.
Fontaine, H., 2003. Âge de conducteurs de voiture et accidents de la route. Quel risque pour les seniors? Recherce Transp. Sécurité 79, 107–120.
Forsyth, E., Maycock, G., Sexton, B. (1995). Cohort study of learner and novice drivers: Part 3, accidents, offences and driving experience in the first three years of driving. Project Report 111. Crowthorne, Berkshire, Transport Research Laboratory.
Hakamies-Blomqvist, L., Raitanen, T., O'Neill, D., 2002. Driver ageing does not cause higher accident rates per km. Transp. Res. Part F 5, 271–274.
Hanson, T.R., Hildebrand, E.D., 2011. Are rural older drivers subject to low-mileage bias? Accid. Anal. Prev. 43, 1872–1877.
Hauer, E., 1995. On exposure and accident rate. Traffic Eng. Control 36, 134–138.
Hauer, E., 2010. Cause, effect and regression in road safety: A case study. Accid. Anal. Prev. 42, 1128–1135.
Hua, P., Charlton, J.J., Koppel, S., Griffiths, D., St Louis, R.M., Di Stefano, M., Darzins, P., Odell, M., Porter, M.M., Myers, A., Marshall, S., 2018. Characteristics of low and high mileage drivers: Findings from the Ozcandrive older driver cohort study. J. Austr. College Road Saf. 29, 53–62.
Janke, M.K., 1991. Accidents, mileage, and the exaggeration of risk. Accid. Anal. Prev. 23, 183–188.
Keall, M.D., Frith, W.J., 2006. Characteristics and risks of drivers with low annual distance driven. Traffic Inj. Prev. 7, 248–255.
Langford, J., Methorst, R., Hakamies-Blomqvist, L., 2006. Older drivers do not have a high crash risk: a replication of low mileage bias. Accid. Anal. Prev. 38, 574–578.
Langford, J., Koppel, S., McCarthy, D., Srinivasan, S., 2008. In defence of the 'low-mileage bias'. Accid. Anal. Prev. 40, 1996–1999.
Langford, J., Charlton, J.L., Koppel, S., Myers, A., Tuokko, H., Marshall, S., Man-Son-Hing, M., Darzins, P., Di Stefano, M., Macdonald, W., 2013. Findings from the Candrive/Ozcandrive study: low mileage older drivers, crash risk and reduced fitness to drive. Accid. Anal. Prev. 61, 304–310.
Lourens, P.F., Vissers, J.A.M.M., Jessurun, M., 1999. Annual mileage, driving violations, and accident involvement in relation to drivers' sex, age, and level of education. Accid. Anal. Prev. 31, 593–597.
Massie, D.L., Green, P.E., Campbell, K.L., 1997. Crash involvement rates by driver gender, age and the role of average annual mileage. Accid. Anal. Prev. 29, 675–685.
Maycock, G., Lockwood, C.R., 1993. The accident liability of British car drivers. Transp. Rev. 13, 231–245.
Mitroff, S.R., Biggs, A.T., 2014. The ultra-rare-item effect: Visual search for exceedingly rare items is highly susceptible to error. Psychol. Sci. 25, 284–289.
Staplin, L., Gish, K.W., Joyce, J., 2008. 'Low mileage bias' and related policy implications – A cautionary note. Accid. Anal. Prev. 40, 1249–1252.
Sterne, J.A.C., Egger, M., 2001. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. J. Clin. Epidemiol. 54, 1046–1055.
Wåhlberg, A. E. af. (2009). Driver behaviour and accident research methodology. Boca Raton, CRC Press. Taylor and Francis Group.
af Wåhlberg, A.E., Dorn, L., 2015. How reliable are self-report measures of mileage, violations and crashes? Saf. Sci. 76, 67–73.