The final publication is available in: Accident Analysis and Prevention, 59, 2013, 301-308. 10.1016/j.aap.2013.06.025

Can it be true that most drivers are safer than the average driver?

Rune Elvik

Institute of Transport Economics

Gaustadalleen 21, NO-0349 Oslo, Norway

E-mail: re@toi.no

ABSTRACT

Surveys finding that a large majority of drivers regard themselves as safer than the average driver have been ridiculed as showing that most drivers are overconfident about their safety and as showing something which is logically impossible, since in a normal distribution exactly half are below average and half above. This paper shows that this criticism is misplaced. Driver accident involvement does not follow a normal distribution, and it is mathematically entirely possible that a huge majority of drivers could be safer than the average driver. The distribution of accidents in a population of drivers is typically skewed, with a majority of drivers not reporting involvement in any accident in the period covered by the data, often a period of 1-3 years. In this paper, examples are given of data sets in which the percentage of drivers who are safer than the average driver ranges from about 70 to 90 percent. The paper explains how, based on knowing the mean and variance of the distribution of accidents in a population of drivers in a given period, the long-term expected number of accidents for drivers who were involved in 0, 1, 2, or more accidents can be estimated. Such estimates invariably show that the huge majority of drivers are safer than the average driver.

Key words: driver safety, optimism bias, regression-to-the-mean, long-term expected number of accidents, negative binomial distribution

1 INTRODUCTION

In 1981 Ola Svenson published a much-quoted paper entitled: “Are we all less risky and more skillful than our fellow drivers?” (Svenson 1981). In the paper, he asked students in the United States and Sweden to place themselves in percentiles with respect to driving skill and safety. Ten percentiles (0-10, 11-20, etc) were listed. The percentiles were ranked so that the first (0-10) indicated the bottom ten percent with respect to skill and safety and the last (91-100) the upper ten percent with respect to skill and safety. If students had a realistic perception of their skill and safety, then, by definition, each percentile should contain ten percent of the students. However, Svenson found that 87.5 percent of US students and 77.1 percent of Swedish students rated their safety in the upper five percentiles, i.e. safer than the median (50^{th} percentile) driver.

Similar results have been reproduced in subsequent studies (see, for example, Svenson, Fishhoff and MacGregor 1985, DeJoy 1989, Holland 1993, Harré and Sibley 2007) and interpreted as showing “optimism bias”. Some authors seem to assume that it is mathematically impossible for more than half of drivers to be safer than average. Thus, Svenson, Fishhoff and MacGregor state (1985, page 119): “Of course, it is no more possible for most people to be safer than average than it is for most to have above average intelligence.” Hence, when more than 50 percent of drivers state that they are safer than the average driver this is interpreted as showing a biased perception of driver safety. It is, however, entirely possible that more than 50 percent of drivers actually are safer than the average driver. The aim of this paper is to show that in representative samples of drivers, it will typically be the case that:

1. A large majority of drivers will not report involvement in an accident during a period of a few years (like 1-3 years).
2. The long-term expected number of accidents for drivers who reported involvement in 0, 1, 2, or more accidents can be estimated if the mean number of accidents per driver and its variance in the population of drivers are known.
3. It will typically be the case that a huge majority of drivers have a long-term expected number of accidents which is slightly lower than the overall mean number of accidents per driver whereas a small minority of drivers have a long-term expected number of accidents which is considerably higher than the overall mean number of accidents per driver.

Obviously, these facts about how accidents are typically distributed in a population of drivers do not imply that drivers who state that they are safer than average are correct in this assessment. However, the characteristics that are common to the distribution of accidents in a population of drivers suggest that it is not necessarily meaningless or impossible, for a majority of drivers to be safer than the average driver. Before showing examples of data sets where the majority of drivers are safer than the average driver, some key concepts of the study will be briefly discussed.

2 KEY CONCEPTS

There are three key concepts in this study that require brief comments: driver safety, average driver, and population of drivers. The safety of a driver is defined as the long-term expected number of accidents per unit of time for that driver. This definition of safety is identical to the definition proposed by Hauer (1997, page 24) stating that safety is "the number of accidents by kind and severity, expected to occur on an entity during a specific period:" An "entity" is any study unit, such as a driver, an intersection, a road section or a vehicle. The expected number of accidents denotes the expected value of a random variable. This definition is preferred to defining safety in terms of an accident rate, i.e. the number of accidents per unit of exposure (e.g. per kilometre of driving), since accident rates tend to be highly non-linear and their values are therefore not proportional to the number of accidents. The relationship between the expected number of accidents and accident rate as estimators of driver safety will be discussed later in the paper.

The concept of an "average driver" can be elaborated in many ways. One might think of an average driver as a driver with an average amount of experience, who drives an average distance, and so on. By enumerating characteristics for which it makes sense to speak about average values, one can make the concept of an average driver very precise. Ultimately, however, defining an average driver as a driver who has average values on a number of variables becomes absurd. What is an average place of residence? Or average gender? The focus of this paper is the distribution of accidents in a population of drivers. When studying the distribution of accidents, individual characteristics of each driver are of no interest. Hence, in this paper an average driver is defined as a driver whose long-term expected number of accidents is equal to the mean number of accidents per driver in the population of drivers to which the driver belongs.

A population of drivers is simply all drivers who are identified in a formal record, such as the record of driving licence holders in a country (or part of a country). It should be possible to enumerate members of the population. In some cases, sub-populations satisfying certain conditions can be formed (e.g. female driver between 18 and 24 years of age). The most commonly studied population of drivers is all driving licence holders in a jurisdiction, but all professional drivers employed by a company may also be regarded as a well-defined population of drivers.

3 THE DISTRIBUTION OF ACCIDENTS AMONG DRIVERS AND THE ESTIMATION OF THE LONG-TERM EXPECTED NUMBER OF ACCIDENTS PER DRIVER

In order to evaluate whether it is common for the majority of drivers to be safer than the average driver, data for a set of populations of drivers have been reviewed:

1. Drivers in Connecticut, USA, with accident data for 1931-36 (Forbes 1939).
2. Bus drivers in Northern Ireland, with accident data for 1952-55 (Cresswell and Froggatt 1963).
3. Drivers in California, USA, with accident data for 1962-68 (Burg 1970).
4. Drivers in California, USA, with accident data from 1961-63 (Weber 1972).
5. Drivers in North Carolina, USA, with accident data for four years (years not stated) (Hauer and Persaud 1983).
6. Drivers in Ontario, Canada, with data from 1981-84 (Hauer, Persaud, Smiley and Duncan 1991).

7. Young drivers in Norway, with data for 1998-99 (Sagberg 2000).

These data sets have been selected because they all permit an evaluation of the accuracy of estimates of the long-term expected number of accidents per driver. The data sets are therefore informative about the distribution of drivers according to the expected number of accidents. Moreover, the data sets cover a long period of time, originate in different countries and include both ordinary drivers and professional drivers.

One of the first researchers who presented data on the distribution of accidents in a population of drivers was Forbes (1939), who in 1939 presented data on the distribution of accidents among 29,531 drivers in the state of Connecticut for two three-year periods: 1931-33 and 1934-36. Forbes states that the data were taken from a report of the Bureau of Public Roads. Table 1 reproduces these data.

Table 1 about here

It is seen that most drivers were not involved in an accident in the first three years. On the average, these drivers were involved in 0.101 accidents in the second three years. 2874 drivers were involved in one accident in the first three years. On the average, these drivers were involved in 0.199 accidents in the second three years. The mean number of accidents per driver among drivers who were involved in 1, 2, 3 or 4 accidents during the first three years was substantially reduced in the second three years. Conversely, the mean number of accidents per driver increased from the first to the second three years among drivers who were not involved in accidents during the first three years.

These changes are an example of regression-to-the-mean. The recorded number of accidents during the first three years is not an unbiased estimator of the long-term expected number of accidents per driver, but is partly the result of random fluctuations around the mean value. The relative contributions from random and systematic variation in the number of accidents to the recorded number of accidents per driver during the first three years can be determined by examining the ratio of the mean to the variance. In general (Hauer 1986):

Variation in the recorded number of accidents = Random variation + Systematic variation

It is generally assumed that random variation in the number of accidents can be modelled statistically by means of the Poisson distribution (Fridstrøm et al. 1995). If variation in the recorded number of accidents was purely random, the variance of the distribution of accidents among drivers would equal the mean number of accidents per driver, because the variance equals the mean in a Poisson-distribution. In the population of drivers in Connecticut during the period 1931-33, the mean number of accidents per driver was 0.126. The variance was 0.145. Hence, $(0.126/0.145) \cdot 100 =$ 86.9 percent of the variation in the recorded number of accidents was random and $[(0.145 - 0.126)/0.145] \cdot 100 = 13.1$ percent was systematic.

The systematic variation in the number of accidents, which is variation in the long-term expected number of accidents per driver, is reflected in the differences in the mean number of accidents per driver in the period 1934-36 for drivers who were involved in 0, 1, 2, 3 or 4 accidents in the period 1931-33. Random variation is eliminated from the first to the second three year period and only systematic variation remains. Thus, the mean number of accidents per driver during the period 1934-36 reflects the long-term expected number of accidents per driver.

The long-term expected number of accidents per driver can be estimated on the basis of knowledge of the mean and variance of the distribution of accidents between drivers in the first period. A simple, but often highly precise estimate of the long-term expected number of accidents is obtained as (Hauer 1986, equation 6):

$$E(m|x) = x + [\left(\frac{E(x)}{Var(x)}\right) \cdot (E(x) - x)] \quad (1)$$

Equation 1 gives an estimate of the long-term expected number of accidents for drivers who have been involved in x accident (x = 0, 1, 2, ..., n). E(x) is the mean number of accidents per driver (0.126 for the Connecticut drivers during 1931-33). Var(x) is the variance of the distribution of accidents between drivers (0.145 for the Connecticut drivers during 1931-33). Thus, for Connecticut drivers who were not involved in an accident during 1931-33, the estimate of the long-term expected number of accidents becomes:

Long-term expected number of accidents = 0 + [(0.126/0.145) · (0.126 – 0)] = 0.110.

While considerably more sophisticated methods for estimating the expected number of accidents have been developed after 1986, the attraction of equation 1 is precisely its simplicity. All you need to know to apply it is the mean number of accidents per driver and the variance of the number of accidents in a population of drivers. As can be seen from Table 1, the predictions of the long-term expected number of accidents obtained by means of equation 1 are quite close to the observed mean values for the years 1934-36. Two versions of the predictions are shown in Table 1. The first is based on data for the period 1931-33 exclusively. The second version has adjusted all predictions by multiplying them with the ratio of the mean number of accidents per driver during 1934-36 (0.114) to the mean during 1931-33 (0.126) to account for the fact that the mean number of accidents per driver was reduced from the first to the second period.

Table 2 presents similar predictions in seven other data sets that permit an assessment of the accuracy of predictions of the long-term expected number of accidents per driver. The data sets presented in table 2 differ with respect to the number of drivers included. By far the largest data set is the North Carolina driver record (Hauer and Persaud 1983), which refers to two periods of two years.

Table 2 about here

The predictions of the long-term expected number of accidents per driver are, in general, very good. A tendency can be seen for the predictions to be more precise in large populations of drivers than in small populations. Figure 1 illustrates the predictions and actual long-term mean values for the largest of the data sets, the North Carolina driver record. The predictions are linear, but very close to the actual mean values for the second period. The predictions are indicated by the data points 0.104, 0.250, ... , 0.979. The actual values are indicated by the data points 0.117, 0.216, ... , 0.944. The correlation between the actual and predicted values is 0.996.

Figure 1 about here

Based on these results, it is concluded that the long-term expected number of accidents per driver can be reliably estimated by applying equation 1. This predictive equation has therefore also been applied to a data set collected by Glad (1988), who only had data for one period for three samples of drivers during the first years of their driving careers. This data set is particularly interesting, because it refers to drivers whose mean annual number of accidents is higher than in the samples of ordinary drivers included in Tables 1 and 2. One can imagine that when the mean number of accidents per driver is higher, the distribution is less skewed and fewer drivers will be safer than the average driver. Table 3 presents the distribution of accidents in the three samples studied by Glad (1988).

Table 3 about here

The majority of drivers in all three samples were not involved in any accident. These drivers are safer than the average driver in the sample they belong to. Thus, even among drivers who are early in their driving careers, the majority tend to be safer than the average driver at a similar stage of the driving career.

4 HOW MUCH SAFER THAN THE AVERAGE DRIVER IS THE MAJORITY OF DRIVERS?

It remains to determine how much safer than the average driver the majority of drivers are, and how large this majority of drivers is. Figure 2 sheds light on this question for drivers in North Carolina (Hauer and Persaud 1983). The overall mean number of accidents is shown by the horizontal line located at the value of 0.130 (the mean number of accidents per driver in the second period). The conditional expected number of accidents for drivers who during the first period were involved in 0, 1, 2, or more accidents is shown by the black line with data points 0.117, 0.216, 0.348, and

0.534. The first three of these values can be found in table 2; the fourth is an average value for drivers who were involved in three or more accidents in the first period.

Figure 2 about here

It is seen that about 90 percent of the drivers are safer than the average driver. These drivers had a long-term mean expected number of accidents of 0.117. The overall mean number of accidents per driver in the second period was 0.130. Thus, the majority of drivers were 10 percent safer than the average driver. A minority of drivers, about 10 percent, were considerably less safe than the average driver.

Data for a total of eleven samples of drivers has been presented in this paper. In all these samples, a huge majority of drivers were safer than the average driver. The share of drivers who were safer than the average driver varies between 63 percent and 90 percent. Figure 3 shows the relationship between the overall mean number of accidents per driver and the share of drivers who are safer than the average driver.

Figure 3 about here

It is seen that the share of drivers who are safer than the average driver declines as the overall mean number of accidents per driver increases. The two rightmost data points in Figure 3 refer to the bus drivers in Northern Ireland, whose mean number of accidents was considerably higher than the mean number of accidents in the general population of drivers. Still, even among these drivers, the majority was safer than the average. The relationship between the overall mean number of accidents per driver and the share of drivers who are safer than the average driver is summarised by the power function fitted to the data points in Figure 3. By extrapolating this function, it can be estimated that when the overall mean number of accidents per driver is greater than 10, only a minority of drivers will be safer than the average driver. However, one should not place much confidence in the extrapolation of any function. If the two data points referring to the professional drivers are omitted, a linear function best fits the data. This function predicts that a minority of drivers will be safer than average when the mean number of accidents exceeds 0.76 – a prediction which would have been grossly erroneous for the bus drivers in Northern Ireland.

In the data sets included in this paper, it is typically the case that a large majority of drivers are slightly safer than the average driver. The relationship between the overall mean number of accidents per driver and the percentage difference in the long-term expected number of accidents between the drivers who are safer than the average driver and the average driver is shown in Figure 4.

Figure 4 about here

There is no clear relationship between the population mean number of accidents and how much safer than this average the majority of drivers are. In most samples, however, the difference is quite small. The safer drivers are typically between 10 and 30 percent safer than the average driver.

5 DISCUSSION

The data presented in this paper show that it is mathematically entirely possible that a huge majority of drivers, indeed close to 90 percent, could be safer than the average driver. One should therefore not dismiss it as nonsensical when a majority of drivers state that they are safer than the average driver. They could very well be correct in this assessment.

The findings in the data sets presented in this paper are very consistent. Yet, these data also have significant limitations. In the first place, the number of data sets was limited. To establish the generality of the findings, it would have been useful to examine a larger number of data sets from different jurisdictions. The data sets included only four countries: Canada, Northern Ireland, Norway and the United states. On the other hand, the data span a considerable period of time, ranging from the nineteen-thirties to the nineteen-nineties. The main findings are very stable over time. Moreover, official driver records, like those for California and North Carolina are representative of all drivers. Samples of drivers obtained in surveys are not necessarily representative of drivers in general, but may be biased by low response rates and lack of reliability in self-reported accident data. However, the data for the bus drivers in Northern Ireland are likely to be complete, or nearly so, as it would be difficult for professional drivers not to report accidents.

In the second place, the safety of drivers is measured in terms of the number of accidents per driver per unit of time. It is rather more common to measure driver safety in terms of accident rate, i.e. the number of accidents per kilometre driven. Driver accident rates are, however, highly non-linear and therefore difficult to interpret. Elvik, Erke and Christensen (2009) comment on the use of accident rates as estimators of safety in the following terms (page 26):

"There are two problems in using accident rates ... in order to control for the effects of differences in exposure on the number of accidents. The first problem arises from the fact that accident rate is not independent of exposure, but tends to decline as exposure increases. This tendency is most clearly evident in driver accident rates, as shown in recent studies (Hakamies-Blomqvist et al. 2002, Fontaine 2003, Langford et al. 2006, Alvarez and Fierro 2008). Thus in the study of Hakamies-Blomqvist et al. (2002), accident rates for drivers aged 26-40 years were:

- 72.4 accidents per million km of driving for drivers whose mean annual driving distance was 1272 km;
- 14.7 accidents per million km of driving for drivers whose mean annual driving distance was 8497 km;
- 5.8 accidents per million km of driving for drivers whose mean annual driving distance was 25536 km.

These accident rates cannot be interpreted as estimates of the probability of accidents. The probability of becoming involved in an accident is not even positively related to the accident rates. The mean annual expected number of accidents can be estimated to 0.092 for low-mileage drivers, 0.125 for middle-mileage drivers and 0.148 for high-mileage drivers (estimated by multiplying accident rate by annual mileage). If the assumption is made that accidents occur according to the Poisson probability law, the probability of becoming involved in at least one accident during a year can be estimated to:

- 0.088 for drivers who drive a mean annual distance of 1272 km;
- 0.117 for drivers who drive a mean annual distance of 8497 km;
- 0.138 for drivers who drive a mean annual distance of 25536 km.

In other words: as exposure increases, so does the probability of becoming involved in an accident, but the each additional kilometre driven becomes safer."

Despite these problems in using accident rates, it is fair to ask if the same results would have been obtained if driver accident involvement had been measured as the rate of accidents per kilometre driven. There is, unfortunately, very limited evidence on this question. The most comprehensive estimates of driver accident rates are those that are based on national household travel behaviour surveys. A recent Norwegian study (Bjørnskau 2011) found that drivers with a lower than average rate of involvement in injury accidents accounted for about 71 percent of all kilometres driven. However, since these drivers drive longer annual distances than other drivers, they would be expected to constitute a lower share of drivers than of kilometres driven. The record of driving licence holders shows that the safer drivers make up nearly 68 percent of all licence holders. Thus, even if accident rates are used to estimate driver safety, it could be the case that the majority of drivers are safer than the average for all drivers.

The limitations of the study suggest that one should not conclude that the findings reported in this paper apply to any population of drivers and any indicator of accident involvement. It is likely that the findings have a more limited generality. However, the main objective of this paper was not to establish a finding that applies to any group of drivers. The objective was much more modest: it was simply to show that it is mathematically possible for a majority of drivers to be safer than the average driver.

6 CONCLUSIONS

The main conclusions of the study presented in this paper can be summarised as follows:

1. It is mathematically entirely possible that a large majority of driver can be safer than the average driver. Examples of data sets where this is true have been given.
2. In a population of ordinary drivers, the mean annual number of accidents will often be in the range of 0.05 to 0.20 accidents per driver. It is then highly likely that a large majority of drivers will be safer than the average driver.
3. The findings of this paper apply to the number of accidents per driver per year. They may not necessarily apply when driver accident involvement is stated as the rate of accidents per kilometre of driving.

ACKNOWLEDGEMENT

The study presented in this paper was funded by the Research Council of Norway through grant number 208437.

REFERENCES

Alvarez, F. J., Fierro, I. 2008. Older drivers, medical condition, medical impairment and crash risk. Accident Analysis and Prevention, 40, 55-60.

Bjørnskau, T. 2011. Risiko i veitrafikken 2009-2010. Rapport 1164. Oslo, Transportøkonomisk institutt.

Burg, A. 1970. The stability of driving record over time. Accident Analysis and Prevention, 2, 57-65.

Cresswell, W. L., Froggatt, P. 1963. The causation of bus driver accidents. An epidemiological study. Published for the Nuffield Provincial Hospital Trust by Oxford University Press. London.

DeJoy, D. M. 1989. The optimism bias and traffic accident risk perception. Accident Analysis and Prevention, 21, 333-340.

Elvik, R., Erke, A., Christensen, P. 2009. Elementary units of exposure. Transportation Research Record, 2103, 25-31.

Fontaine, H. 2003. Driver age and road accidents. What is the risk for seniors? Recherche Transports Sécurité, 79 107-120.

Forbes, T. W. 1939. The normal automobile driver as a traffic problem. Journal of General Psychology, 20, 471-474.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., Krogsgård Thomsen, L. 1995. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. Accident Analysis and Prevention, 27, 1-20.

Glad, A. 1988. Fase 2 i føreropplæringen. Effekt på ulykkesrisikoen. Rapport 0015. Oslo, Transportøkonomisk institutt.

Hakamies-Blomqvist, L., Raitanen T., O'Neill, D. 2002. Driver ageing does not cause higher accident rates per km. Transportation Research Part F, 5, 271-274.

Harré, N., Sibley, C. G. 2007. Explicit and implicit self-enhancement biases in drivers and their relationship to driving violations and crash-risk optimism. Accident Analysis and Prevention, 39, 1155-1161.

Hauer, E. 1986. On the estimation of the expected number of accidents. Accident Analysis and Prevention, 18, 1-12.

Hauer, E. 1997. Observational before-after studies in road safety. Estimating the effect of highway and traffic engineering measures on road safety. Oxford, Pergamon Press.

Hauer, E., Persaud, B. 1983. Common bias in before-and-after accident comparisons and its elimination. Transportation Research Record, 905, 164-174.

Hauer, E., Persaud, B. N., Smiley, A., Duncan, D. 1991. Estimating the accident potential of an Ontario driver. Accident Analysis and Prevention, 23, 133-152.

Holland, C. A. 1993. Self-bias in older drivers' judgments of accident likelihood. Accident Analysis and Prevention, 25, 431-441.

Langford, J., Methorst, R., Hakamies-Blomqvist, L. 2006. Older drivers do not have a high crash risk – a replication of low mileage bias. Accident Analysis and Prevention, 38, 574-578.

Sagberg, F. 2000. Evaluering av 16-årsgrense for øvelseskjøring med personbil. Ulykkesrisiko etter førerprøven. Rapport 498. Oslo, Transportøkonomisk institutt.

Svenson, O. 1981. Are we all less risky and more skillful than our fellow drivers? Acta Psychologica, 47, 143-148.

Svenson, O. Fischhoff, B., MacGregor, D. 1985. Perceived driving safety and seatbelt usage. Accident Analysis and Prevention, 17, 119-133.

Weber, D. C. 1972. An analysis of the California driver record study in the context of a classical accident model. Accident Analysis and Prevention, 4, 109-116.

LIST OF FIGURES AND TABLES

Figure 1:

Accuracy of predicted long-term expected number of accidents per driver – based on North Carolina driver record

Figure 2:

Overall and conditional mean long-term expected number of accidents per driver for North Carolina drivers

Figure 3:

Relationship between overall mean number of accidents per driver and the proportion of drivers who are safer than the average driver

Figure 4:

Relationship between overall mean number of accidents per driver and how much safer than the average driver the majority of drivers are

Table 1:

Distribution of accidents among drivers in Connecticut, USA, 1931-33 and 1934-36. Based on Forbes (1939)

Table 2:

Prediction of accidents in data sets showing distribution of accidents in two periods in populations of drivers

Table 3:

Recorded and predicted long-term expected number of accidents for three samples of drivers during the first years of their driving careers. Based on Glad (1988)

Table 1: Distribution of accidents among drivers in Connecticut, USA, 1931-33 and 1934-36. Based on Forbes (1939)

Number of accidents 1931-33	Number of drivers	Mean number of accidents 1934-36	Predicted number of accidents 1934-36 (1)	Predicted number of accidents 1934-36 (2)
0	26259	0.101	0.110	0.099
1	2874	0.199	0.239	0.216
2	357	0.300	0.368	0.333
3	31	0.484	0.497	0.449
4	10	0.700	0.626	0.566
Mean or total	29531	0.114	0.126	0.114
Mean 1931-33	0.126			
Variance 1931-33	0.145			

(1) Prediction based on the mean number of accidents per driver during 1931-33 (0.126)

(2) Prediction adjusted for the change in the mean number of accidents per driver from 1931-33 to 1934-36 (0.114/0.126)

Table 2: Prediction of accidents in data sets showing distribution of accidents in two periods in populations of drivers

Number of accidents first period	Number of drivers	Mean number of accidents second period	Predicted number of accidents second period (1)	Predicted number of accidents second period (2)
Ulster Transport Authority bus drivers (largest group) (Cresswell and Froggatt 1963)				
0	224	0.768	1.044	0.810
1	226	0.951	1.235	0.958
2	150	1.073	1.427	1.107
3	68	1.250	1.618	1.255
4+	40	1.900	1.925	1.493
Mean or total	708	1.001	1.291	1.001
Belfast Corporation Transport trolley bus drivers (Cresswell and Froggatt 1963)				
0	40	1.400	1.391	1.299
1	52	1.962	1.786	1.668
2	62	1.806	2.181	2.037
3	36	2.222	2.576	2.406
4	24	3.500	2.971	2.775
5+	30	3.000	3.773	3.523
Mean or total	244	2.148	2.299	2.148
Data from California driver record study (Burg 1970)				
0	6285	0.230	0.201	0.214
1	1248	0.356	0.387	0.411
2	248	0.500	0.572	0.608
3	51	0.627	0.758	0.806
4	8	0.500	0.944	1.003
5	1	5.000	1.129	1.200
Mean or total	7841	0.262	0.247	0.262
Data from California driver record study (Weber 1972)				
0	129660	0.057	0.055	N.A.
1	16305	0.099	0.106	N.A.
2	1967	0.133	0.157	N.A.
3	212	0.184	0.208	N.A.
4+	38	0.421	0.273	N.A.
Mean or total	148182	0.072	0.072	N.A.

Table 2, continued: Prediction of accidents in data sets showing distribution of accidents in two periods in populations of drivers

Number of accidents first period	Number of drivers	Mean number of accidents second period	Predicted number of accidents second period (1)	Predicted number of accidents second period (2)
Data from North Carolina driver record (Hauer and Persaud 1983)				
0	2234577	0.117	0.104	0.111
1	235080	0.216	0.250	0.267
2	27919	0.348	0.396	0.422
3	3953	0.499	0.540	0.577
4	584	0.703	0.688	0.733
5	99	0.848	0.834	0.888
6	18	0.944	0.979	1.044
Mean or total	2502230	0.130	0.122	0.130
Data from Ontario driver record – males 26-30 years (Hauer et al. 1989)				
0	56656	0.133	0.144	0.131
1	8243	0.199	0.248	0.226
2	963	0.416	0.352	0.320
3	109	0.376	0.455	0.415
4+	26	1.038	0.577	0.527
Mean or total	65997	0.146	0.161	0.146
Data from Norwegian study of novice drivers –age 18-19 (Sagberg 2000)				
0	22881	0.092	0.102	0.088
1	2482	0.192	0.275	0.237
2	260	0.312	0.449	0.387
3	30	0.933	0.623	0.536
4+	13	1.769	1.036	0.892
Mean or total	25666	0.106	0.123	0.106

(1) Prediction based on the mean number of accidents per driver during the first period

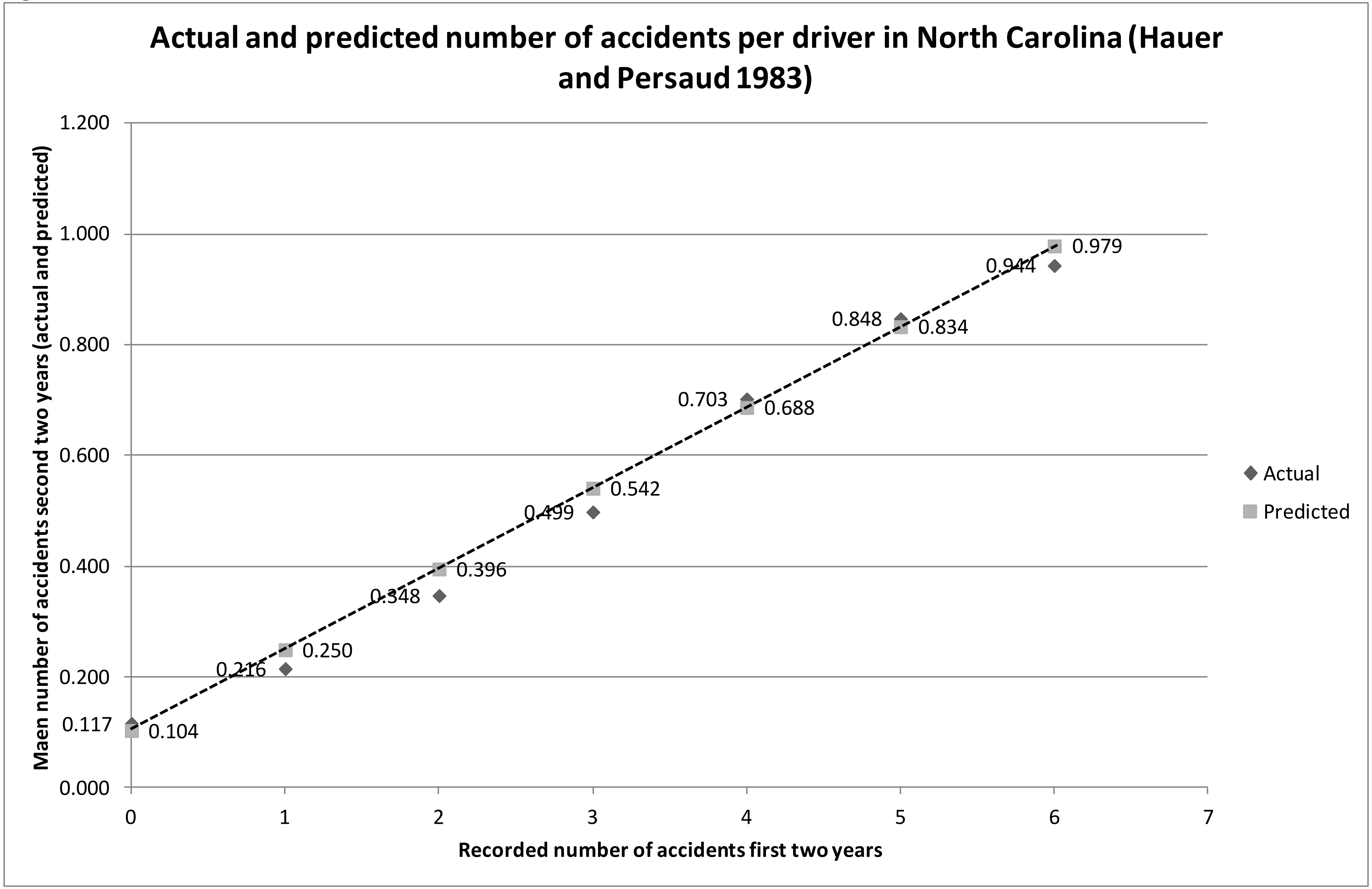
(2) Prediction adjusted for change in the mean number of accidents per driver from the first to the second period

N.A. = not applicable (the two periods were of different length)

Table 3: Recorded and predicted long-term expected number of accidents for three samples of drivers during the first years of their driving careers. Based on Glad (1988)

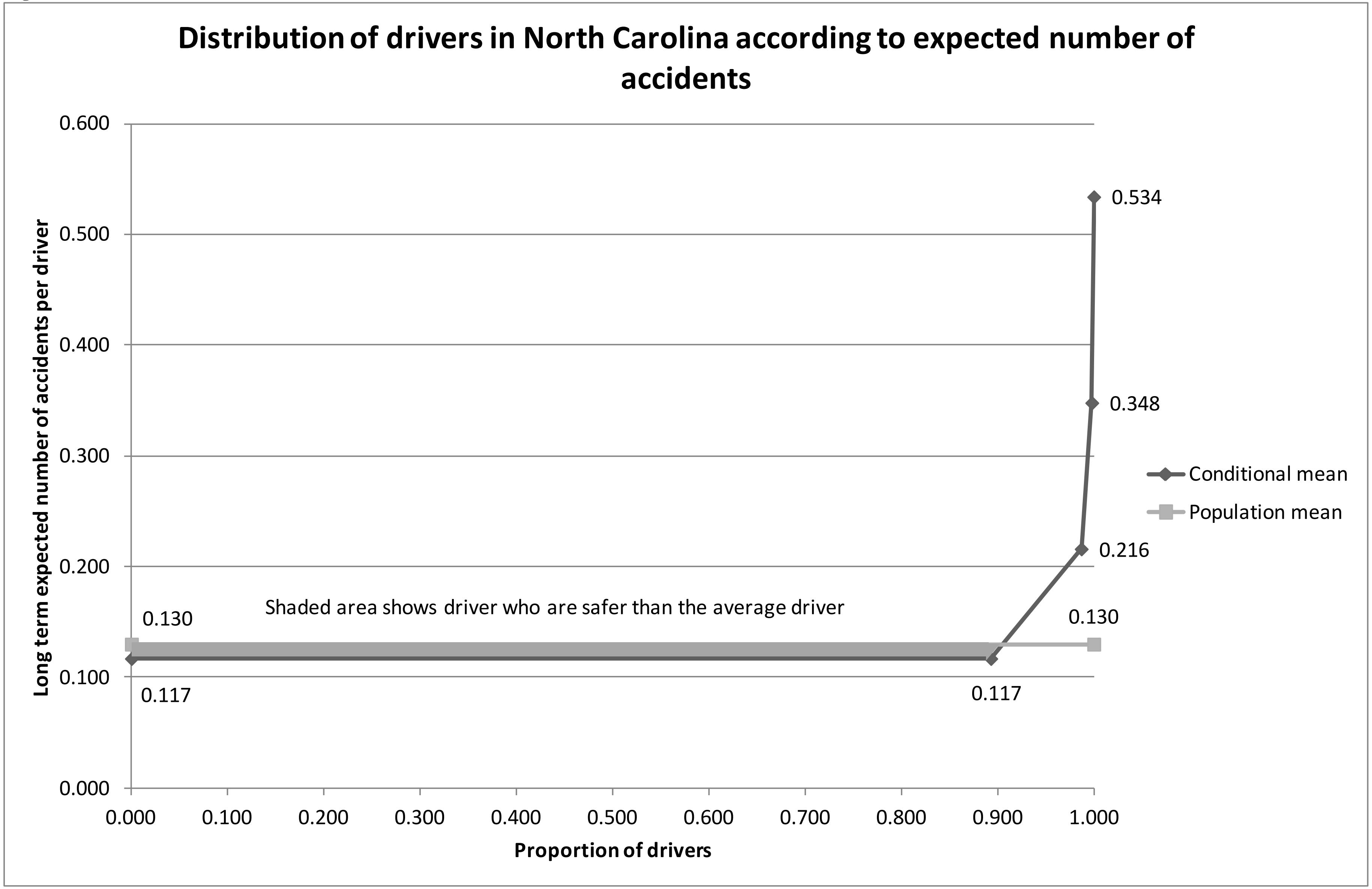
	Drivers with 1-2 years of experience		Drivers with 3-4 years of experience		Drivers with 5-6 years of experience	
Number of accidents	**Number of drivers**	**Long-term expected number of accidents**	**Number of drivers**	**Long-term expected number of accidents**	**Number of drivers**	**Long-term expected number of accidents**
0	2525	0.301	2356	0.312	2633	0.275
1	820	0.615	747	0.569	706	0.464
2	225	0.930	211	0.825	177	0.654
3	55	1.245	55	1.082	37	0.843
4	24	1.560	11	1.339	8	1.033
5	8	1.874	6	1.596	1	1.222
6	2	2.189	0	1.852		
7	1	2.504	2	2.109		
8	2	2.819				
N	3662		3388		3562	
Mean	0.439		0.420		0.339	
Variance	0.640		0.565		0.418	

Figure 1:



Actual and predicted number of accidents per driver in North Carolina (Hauer and Persaud 1983)
1.200
1.000
0.800
0.600
0.400
0.200
0.117
0.000
0.104
0.216
0.250
0.348
0.396
0.499
0.542
0.703
0.688
0.848
0.834
0.944
0.979
0
1
2
3
4
5
6
7
Maen number of accidents second two years (actual and predicted)
Recorded number of accidents first two years
Actual
Predicted

Figure 2:



Distribution of drivers in North Carolina according to expected number of accidents
0.600
0.500
0.400
0.300
0.200
0.100
0.000
Long term expected number of accidents per driver
0.534
0.348
0.216
0.130
0.130
0.117
0.117
Shaded area shows driver who are safer than the average driver
Conditional mean
Population mean
0.000
0.100
0.200
0.300
0.400
0.500
0.600
0.700
0.800
0.900
1.000
Proportion of drivers

Figure 3:

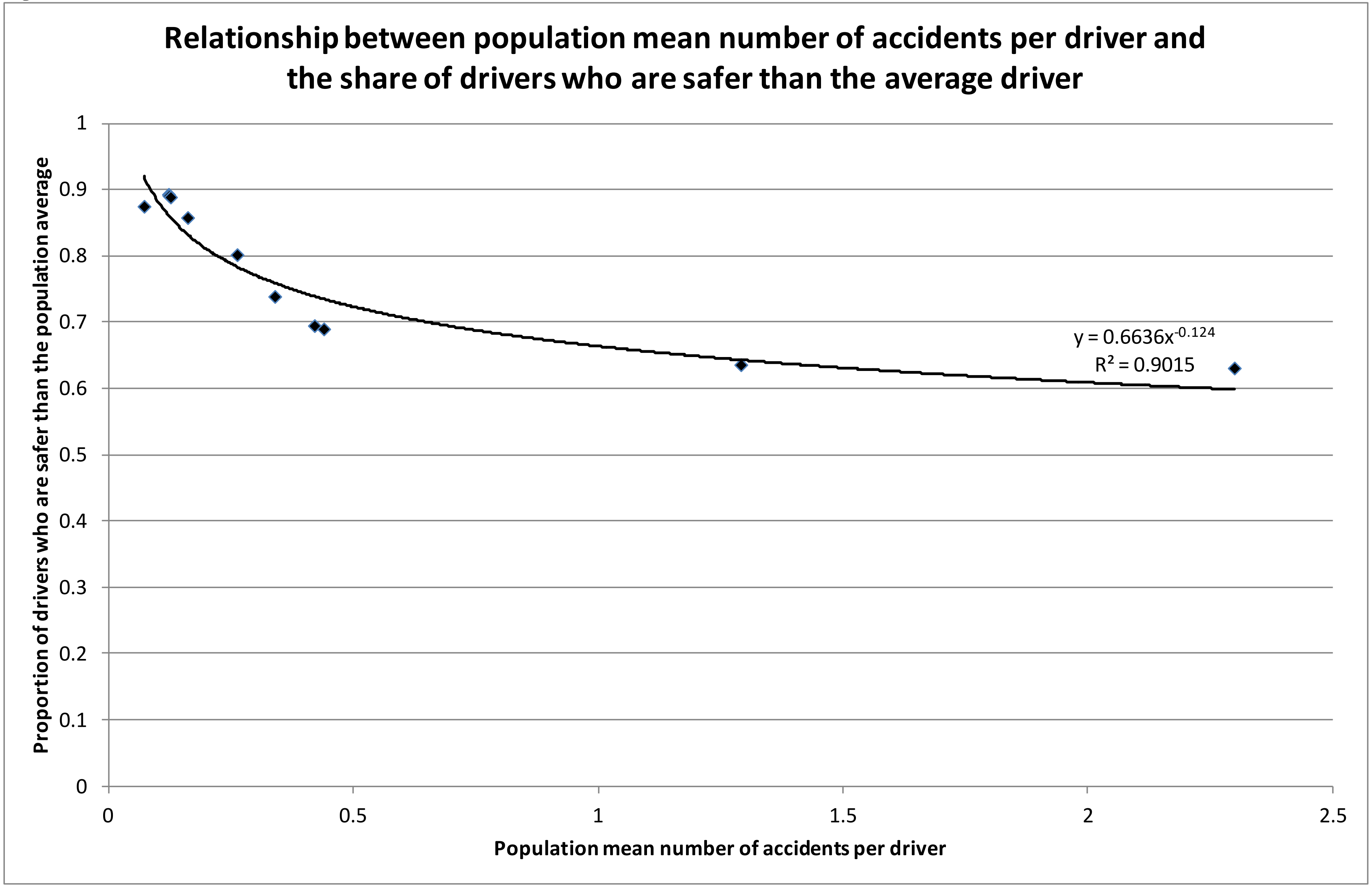
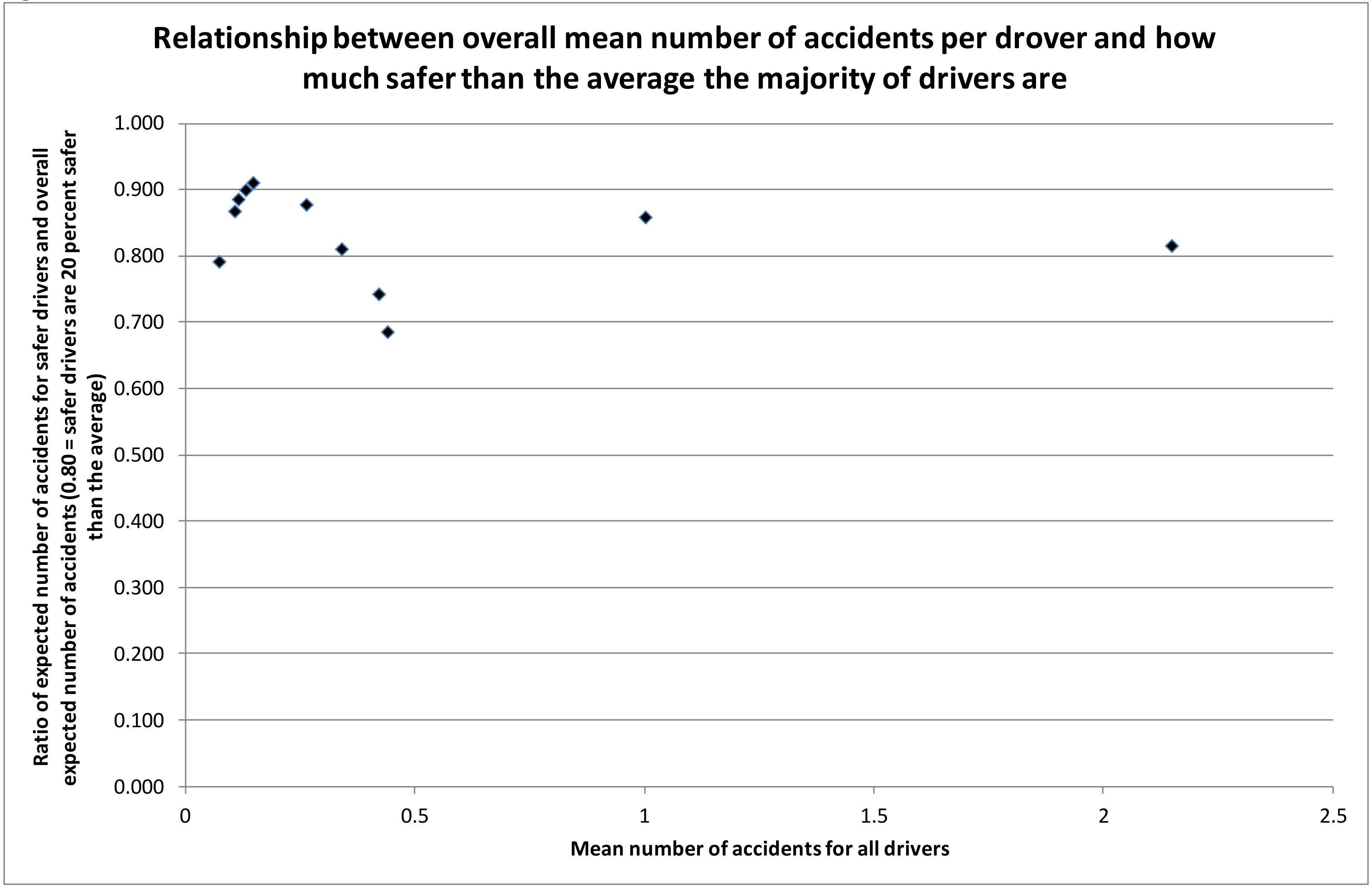

Relationship between population mean number of accidents per driver and the share of drivers who are safer than the average driver
Proportion of drivers who are safer than the population average
Population mean number of accidents per driver
y = 0.6636x-0.124
R² = 0.9015
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.5
1
1.5
2
2.5

Figure 4:


Relationship between overall mean number of accidents per drover and how much safer than the average the majority of drivers are
Ratio of expected number of accidents for safer drivers and overall expected number of accidents (0.80 = safer drivers are 20 percent safer than the average)
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
0
0.5
1
1.5
2
2.5
Mean number of accidents for all drivers