Accident Analysis and Prevention 59 (2013) 301–308

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

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

ARTICLE INFO

Article history:
Received 10 March 2013
Received in revised form 5 June 2013
Accepted 19 June 2013

Keywords:
Driver safety
Optimism bias
Regression-to-the-mean
Long-term expected number of accidents
Negative binomial distribution

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%. 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.

1. Introduction

In 1981 Ola Svenson published a much-quoted paper entitled: “Are we all less risky and more skilful 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 10% with respect to skill and safety and the last (91–100) the upper 10% with respect to skill and safety. If students had a realistic perception of their skill and safety, then, by definition, each percentile should contain 10% of the students. However, Svenson found that 87.5% of US students and 77.1% of Swedish students rated their safety in the upper five percentiles, i.e. safer than the median (50th percentile) driver.

Similar results have been reproduced in subsequent studies (see, for example, Svenson et al., 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% 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% 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.

* Tel.: +47 48943353; fax: +47 22 609200.
E-mail address: re@toi.no

0001-4575/$ – see front matter
http://dx.doi.org/10.1016/j.aap.2013.06.025

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 to 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 to 84 (Hauer et al., 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.

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):

$$\text{Variation in the recorded number of accidents} = \text{Random variation} + \text{Systematic variation}$$

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, $(0.126/0.145) \times 100 = 86.9\%$ of the variation in the recorded number of accidents was random and $[(0.145 - 0.126)/0.145] \times 100 = 13.1\%$ 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