Computing and interpreting accident rates for vehicle types or driver groups.

Ezra Hauer, 35 Merton St., Apt. 1706, Toronto, ON, Canada, M4S 3G4

Tel.: 416 483 4452

Fax: 416 483 4415

e-mail: Ezra.Hauer@utoronto.ca

Word count: 5074
Computing and interpreting accident rates for vehicle types or driver groups.

Ezra Hauer, 35 Merton St., Apt. 1706, Toronto, ON, Canada, M4S 3G4., Ezra.Hauer@utoronto.ca

ABSTRACT

This paper deals with two questions. The first question is which of the many rates that have been proposed for measuring the safety of certain vehicle types or driver groups is to be trusted. We conclude that for a rate to be correct, the numerator and the denominator must pertain to the same entity. If vehicle exposure is in the denominator, then the count of vehicles of some type in accidents (and not of accidents involving vehicles of some type) must be in the numerator; if driver exposure is in the denominator then the count of drivers of some kind in accidents (not of accidents involving that kind of driver) must be in the numerator. The second question is what meaning can be attributed to a finding of over-representation. We conclude that because we always use reported accidents of specified severity, over-representation may be caused by a mix of three factors: the probability to be in an accident per unit of exposure, the probability of the accidents to be reported, and the probability of the accident to be of the specified severity. It follows, that an indication of over-representation cannot be taken to mean that the entity has a larger than normal chance to be involved in accidents nor that one should seek remedies that reduce that chance of the entity to be involved in accidents.

1. INTRODUCTION

There is a rich research literature comparing the safety of various vehicle types and driver groups. The accumulated knowledge serves to make informed decisions about a variety of practical issues. A common question is whether a certain vehicles of a specified vehicle type (or vehicles having certain traits), or drivers of a certain group (drivers with specified characteristics) are less safe than others. To answer the question, the extant research results are reviewed so as to establish what
appears to be the current state of knowledge. In such reviews diverse and sometimes contradictory results are found. The onus is on the reviewer to separate the wheat from the chaff. Bowman and Hummer (1) found themselves in this situation when reviewing the literature about truck accidents on urban freeways. I was facing this predicament recently, when preparing a review for a study on truck size and weight regulation. To do it right I had to clarify to myself what which results are right and which are wrong.

2. CORRECT AND INCORRECT DEFINITIONS OF ACCIDENT RATE

Accident rate is habitually defined as the ratio of accidents and exposure. Traditionally it has been used to describe the safety of a road or the risk to people using it. The same definition of accident rate is at times used to compare the risk of various vehicle types or driver groups. In this case a commonly used definition is:

$$\text{Accident Rate} = \frac{\text{Number of collisions involving a given vehicle type (driver group)}}{\text{Exposure by vehicles (drivers) of the given type (group)}}$$  \hspace{1cm} (1)

Bowman and Hummer (1) point to a pitfall in the accident rate computation for trucks when it is calculated by equation 1.

In their numerical example, trucks have 20% and cars 80% of the total exposure. If it is assumed that cars and trucks are equally likely to be involved in an accident per unit of exposure (the ‘equiprobability assumption’), they should have equal accident rates. Bowman and Hummer show that the use of equation 1 violates this expectation. Under the equiprobability assumption, the proportion of two-vehicle accidents involving two trucks is 0.2×0.2=0.04 and the proportion involving a truck and a car is 2×0.2×0.80=0.32. Thus, while trucks have 20% of exposure they are
expected to be involved in 36% of accidents. Similarly, the proportion of accidents involving cars is $0.8 \times 0.8 + 2 \times 0.2 \times 0.8 = 0.96$ and yet they have 80% of exposure. Obviously there is something wrong with the definition of the accident rate in equation 1, since both trucks and cars appear to be overrepresented in accidents and, in spite of the assumption of equiprobability, their accident rates are different. From this, Bowman and Hummer correctly concluded that it is improper to calculate the accident rate by equation 1.

Their argument rests on the unstated assumption that the probability of one of the vehicles in a two-vehicle accident being a car or a truck does not depend on what the other vehicle in the accident is. This hidden assumption of 'statistical independence' is not likely to hold when the traffic stream is not a random mix of trucks and cars. However, invoking the 'statistical independence' assumption does not detract from the force of their argument. A valid definition of the accident rate should produce unbiased results under all conditions, including that when the outcomes are statistically independent.

The root of the problem with equation 1 is that the numerator pertains to apples and the denominator to oranges. Since the denominator is the exposure of *vehicles*, the numerator should also be the count of *vehicles*, not the count of *accidents*. To illustrate, under the equiprobability and independence assumptions, and with the 20%-80% share of exposure, in 100 two-vehicle accidents we expect the following counts:
TABLE 1 Counts of Accidents and Vehicles

<table>
<thead>
<tr>
<th>Accident type</th>
<th>Number of Accidents</th>
<th>Number of trucks involved</th>
<th>Number of cars involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truck-Truck (TT)</td>
<td>100×0.2×0.2=4</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Truck-Car (TC)</td>
<td>100×0.2×0.8=16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Car-Truck (CT)</td>
<td>100×0.2×0.8=16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Car-Car (CC)</td>
<td>100×0.8×0.8=64</td>
<td>0</td>
<td>128</td>
</tr>
<tr>
<td>Sums</td>
<td>100</td>
<td>40 trucks</td>
<td>160 cars</td>
</tr>
</tbody>
</table>

The count of accidents leads to the problem identified earlier. But the count of vehicles is seen to lead to the correct result. Thus, in 100 accidents of two vehicles we have the truck proportion of 40 trucks/200 vehicles=0.2 and car proportion of 160 cars/200 vehicles=0.8, in correspondence with their share of exposure. It follows that in two-vehicle accidents it is the \((\text{Number of vehicles of a specified type involved in two-vehicle accidents})/(2\times\text{Number of two-vehicle accidents})\) that should correspond to the proportion of exposure by the vehicles of the specified type.

The same logic applies to 3, 4, 5,. . .,n-vehicle accidents. To show this generally, let \(E_T\) and \(E_C\) denote truck and car exposures. Define the share of truck exposure by:

\[
P = \frac{E_T}{E_T + E_C}
\]

Now 1-P is the share of car of exposure. When trucks and cars are equally likely to be involved in an accident, P is also the probability that a vehicle in a multivehicle accident is a truck. Let \(N_n\) be the number accidents involving \(n\) vehicles. Under the statistical independence assumption, the probability to have ‘i’ trucks amongst the ‘n’ vehicles in an accident is given by the binomial
probability mass function $\binom{n}{i}P^i(1-P)^{n-i}$. The expectation of the binomial distribution is $nP$. Thus the number of trucks expected to be involved in $N_n$ accidents of $n$ vehicles is $nP_n$. By the same argument the number of cars expected to be involved in $N_n$ accidents of $n$ vehicles is $n(1-P)N_n$. It follows that in accidents involving $n$ vehicles the expression $(\text{Number of vehicles of a specified type involved in } n\text{-vehicle accidents})/(n \times \text{Number of } n\text{-vehicle accidents})$ is $nP_n/(nN_n)=P$ for trucks and $n(1-P)N_n/(nN_n)=1-P$ for cars, as required.

A part of the Bowman and Hummer (1) paper is based on the earlier work by Khasnabis and Reddy (2) who call equation 1 the “traditional approach”. Khasnabis and Reddy also recognized that the deficiency of equation 1 resides in the fact that “a truck accident is generally referred to as one that involves at least one truck. By the same token, an accident that involves at least one passenger car is a passenger car accident.” (p. 37) So there is double counting, because some accidents count as both truck accidents and as car accidents. The double counting results in the phantom over-representation bias.

Khasnabis and Reddy consider three alternative approaches to replace the deficient traditional approach. Their ‘Approach 1’ is to divide the number of accidents involving at least one truck by a modified exposure defined as: Truck VMT + Car VMT × (number of cars in truck accidents)/(number of cars in all accidents). This is an ad hoc approach that instead of remedying the double counting in the numerator inflates the denominator. Approach 1 will not produce correct results except by coincidence, and the authors rightly recommend against its use.

To avoid the double counting, Khasnabis and Reddy recommend Approach 3 which is to define ‘Truck-Only’ and ‘Passenger-Car-Only’ accident rates. These have in the numerator the number of accidents involving only trucks and or only passenger cars. In Approach 3, data about
accidents that involve both a truck and a car are simply not used. This is a serious loss of data since most truck accidents are with cars. However, the shortcoming could be overlooked if the approach produced logically coherent results. Unfortunately, ‘Approach 3’ does not meet the basic requirement that if two vehicle types have the same probability to be involved in an accident per unit of exposure, their expected accident rates will be equal. Simulation results show that under conditions of equiprobability and independence the ‘Truck-Only’ and ‘Car-Only’ accident rates are wildly different.

It remains to examine their ‘Approach 2’ which is embodied in equation 3.

\[
Truck\ involvement\ rate = \frac{\text{Number of trucks involved in accidents}}{\text{Total truck VMT}}
\]  

(Khasnabis and Reddy recommend against the use of equation 3 on the ground that “the use of vehicles in the numerator (as opposed to accidents) would inflate the rate for passenger cars due to the simple fact the most multivehicle truck accidents involve passenger cars as the other vehicle whereas most multivehicle passenger-car accidents involve another passenger car.” (p. 38) This is why Approach 2 “was not pursued” by Khasnabis and Reddy. Ten years later, Bowman and Hummer echo the same argument saying that “The use of vehicles in the numerator (as opposed to accidents) inflates the passenger car rate . . .”.

The opinion that use of equation 3 would “overly exaggerate the adverse role of passenger cars in highway accidents in comparison with trucks” (Khasnabis and Reddy, 2, p. 38) is incorrect. As has been shown earlier, under the assumption of independence and the assumption that per unit of exposure cars and trucks are equally likely to be in accidents, the expected number of trucks
involved in accidents of \( n \) vehicles is \( nPN_n \). Let \( N \) be the sum of \( N_n \). With this notation, the expected number of trucks involved in all \( N \) accidents is:

\[
\sum_{n=1}^{\infty} nPN_n = PN \sum_{n=1}^{\infty} \frac{N_n}{N} = P \times (\text{Number of collisions}) \times (\text{Average number of vehicles in a collision})
\]

A parallel argument can be used to show that the expected number of cars involved in \( N \) accidents is \((1-P) \times (\text{Number of accidents}) \times (\text{Average number of vehicles in an accident})\). Thus, the ratio of the expected number of cars in accidents to expected number of trucks in accidents is \((1-P)/P\). Using now the definition of \( P \) in equation 2,

\[
\frac{E_C}{E_T} = \frac{1-P}{P} \times \frac{E_C}{E_T} = \frac{E_C}{E_T} = \text{car exposure} / \text{truck exposure}
\]

other words, if cars and trucks have in fact the same probability of being in accidents per unit of exposure, and when the independence conditions holds, then the expected involvement rate of trucks and cars as given by equation 3 would be the same, as required.

Thus, at least when the assumptions of independence and equiprobability hold, equation 3 avoids the pitfall of equation 1. To check whether equation 3 will pass muster when the assumption of independence is relaxed a simulation program has been written. It allows one to specify a variety of traffic conditions which vary from an entirely random mix of trucks and cars to where trucks tend to follow trucks and cars tend to follow cars. With the exposures of 20% trucks and 80% cars we
Ezra Hauer

examined the simulated ratio of cars-in-accidents/trucks-in-accidents. The specified conditions ranged from when trucks tend to follow trucks with a probability of 0.2 (which is equivalent to a random mix) up to a probability of 0.99 (when trucks will almost always collide with trucks and cars with cars). In all cases the simulated ratio was within one percent of the ratio of exposures. While simulation results cannot prove the general verity of an assertion, they provide reasonable assurance that equation 3 gives unbiased estimates even when the independence assumption does not hold.

By the equiprobability assumption cars and trucks have the same probability to be in an accident per unit of exposure. Suppose now that this assumption is untrue. This can be represented by a factor ‘F’ multiplying the exposure of trucks. Thus, if trucks were F times as likely to be in an accident as cars it would be as if their exposure was not $E_T$ but $F \times E_T$ while the equiprobability condition held. With this $P = \frac{F E_T}{F E_T + E_C}$ and the expected ratio of cars and trucks is $\frac{1}{P} = \frac{E_C}{F E_T}$. It follows that F can be estimated by:

$$\text{Estimate of } F = \frac{\text{truck involvement rate}}{\text{car involvement rate}}$$

Both Khasnabis and Reddy and Bowman and Hummer seek the remedy to equation 1 by finding a more appropriate definition of exposure. While the concept of ‘exposure’ is indeed fraught with difficulties, and the accident rate is not a constant but a function of the use of a facility by all vehicles types, the problem that Bowman and Hummer pointed to is in the denominator, not in the numerator. The solution that Khasnabis and Reddy considered and rejected was the right one. If exposure in the denominator pertains to vehicles, one should count vehicles (not accidents) in the numerator. Equation 1 is wrong and equation 3 is correct.
So far we spoke of trucks and cars. More generally, one may wish to compare the accident rates of several vehicle types. Equation 3 remains the correct device. Similarly, instead of examining vehicle safety one may inquire about the safety of drivers in various age and gender groups. Equation 3 remains the proper tool. The numerator should be the count of drivers or vehicle in accidents, not the count of accidents involving a certain vehicle type or driver group.

2. THE ‘WHY-INDETERMINACY’.

Clarke (3) writes that “Medium/heavy trucks ... accumulate 7 percent of all the vehicle miles of travel, while being involved in 8 percent of all fatal crashes...”. Here is how Clarke’s numbers were obtained. In 1995 all U.S. vehicles covered 3.890×10^{14} km (2.423×10^{14} miles) of travel while Medium/heavy trucks covered 0.286×10^{14} km (178×10^{14} miles); 0.286/3.890=0.073 or about 7%. In 1995, 56,524 vehicles were involved in fatal crashes of which 4,472 were Medium/heavy trucks; 4,472/56,525=0.079 or about 8%. Thus, Clarke avoids the pitfall of equation 1, using the correct equation 3.

Does this mean that there is something about medium/heavy trucks and their use (truck instability, long braking distance, high center of gravity, long work hours by truckers, drug use and the like) for which they are more likely to be involved accidents than other vehicles? To answer, it helps to devise a thought experiment.

Assume that many identical vehicles use a road system. Some vehicles have an Amber spot under the hood (the A-vehicles) the rest have a Blue spot under the hood (the B-vehicles). Of course, the color of the spot under the hood has no effect on how accident-prone a vehicle is. The A vehicles have a proportion P of the total exposure and the B-vehicles have the remaining proportion of exposure (1-P). Assume that accidents are proportional to exposure, and that the time or place of
travel or kind of driver are not related to the color of the spot under the hood. Therefore, under the specified assumptions one expects that:

\[ p = \text{expected proportion of A-vehicles in accidents} = \]
\[ = \text{expected proportion of A-vehicles in exposure} = P \]

If under these conditions some method of data analysis is expected to yield the result that \( p \neq P \), the method is unsound. That under the specified assumptions the use of equation 3 leads to \( p = P \) has been shown earlier. Therefore, should some data indicate that \( p > P \) would be led to conclude that the specified conditions do not hold; specifically that the A-vehicles are ‘over-involved’. Over-representation tends to initiates a search for cause - what is it about these vehicles that causes over-representation.

The data we have are always of reported accidents. The question is whether, when trucks and cars are equally likely to be involved in accidents per unit of exposure, the expected equality \( p = P \) will still hold even if analysis is based on reported accidents of a specified severity. The answer to this question is:”No”. That is, even if truck and cars are equally likely to be involved in an accident per unit of travel, their proportion in reported accidents of specified severity will not be the same. To show this, assume again that the identical A and B-vehicles differ in exposure but do not differ in time or place of travel, nor in the kind of driver. However, assume now that the accidents in which the A and B-vehicles are involved differ in the probability to be reported as belonging to some accident severity class. This may be due to differences between A and B vehicles in the inclination to report accidents and also due to differences in the severity of the outcome of the accidents in which these vehicles are involved.
Let $Q_A$ and $Q_B$ be the probabilities for an A and a B one-vehicle accident to be reported in a certain accident severity class. Of $N_1$ one-vehicle accidents occurring we expect to find in the specified severity class a count of $N_1 P Q_A$ accidents involving A-vehicles and a count of $N_1 (1-P) Q_B$ accidents involving B-vehicles. If so,

$$p = \frac{P Q_A}{P Q_A + (1-P) Q_B} = \frac{P}{P + (1-P) \frac{Q_B}{Q_A}} \tag{6}$$

In this case $p=P$ if and only if $Q_A=Q_B$. In all other cases, when there is reason to believe that $Q_A \neq Q_B$, one should expect that $p \neq P$. This is not because of a difference between A and B vehicles in their risk to be in a (single-vehicle) accident; it is entirely due to the difference in the probability of A and B vehicle accidents (once an accident has occurred) to be reported in this severity class.

The same can be generalized to two-vehicle accidents. Let $Q_{AA}$, $Q_{BB}$ and $Q_{AB}$ be the corresponding probabilities of an accident to be reported in a certain severity class for two-vehicle accidents. Of the $N_2 P^2$ accidents involving two A-vehicles we expect $N_2 P^2 Q_{AA}$ accidents to be in the count; of the $N_2 (1-P)^2$ accidents involving two B-vehicles we expect $N_2 (1-P)^2 Q_{BB}$ accidents to be in the count; and of the $2N_2 P(1-P)$ accidents involving an A-vehicle and a B-vehicle we expect $2N_2 P(1-P) Q_{AB}$ accidents to be in the count. Now $p$ is the expression:

$$p = \frac{1 + \frac{1-P}{P} \frac{Q_{AB}}{Q_{AA}}}{1 + 2 \frac{1-P}{P} \frac{Q_{AB}}{Q_{AA}} + \frac{(1-P)^2}{P} \frac{Q_{BB}}{Q_{AA}}} \tag{7}$$

Again, $p=P$ if and only if $Q_{AA}=Q_{AB}=Q_{BB}$. 
The assertion that if A and B vehicles have the same probability to be in an accident per unit of exposure one should expect \( p=P \) now has to be qualified. Obviously \( p \neq P \) except in the rare and unlikely event that all \( Q \)’s are the same. That is, the expectation that \( p=P \) remains correct only if all accidents have the same probability to be reported. When we use data to compare estimates of \( p \) (proportion of vehicles in accidents) and \( P \) (proportion of vehicles in exposure), we usually are interested in a possible over-representation of certain vehicles types, driver classes, or road categories. In such comparisons if the compared vehicles differ in mass, if the drivers differ in vulnerability to injury, if the roads differ in speed, or if there are reasons to suspect differences in the inclination to report an accident, the \( Q \)’s will be different. In such all such circumstances one will find over-representation but it cannot be taken to mean that A and B differ in their probability to be in an accident per unit of exposure.

A numerical example may help to elucidate the implications of this finding. Just as in Clarke (3), we consider the proportion of medium/heavy trucks amongst vehicles in fatal accidents. Let \( P=0.07 \) be their proportion of vehicle miles of travel, as Clarke reports. Consider first single-vehicle accidents. The chance \( (Q_A) \) of a single vehicle accident by a medium/heavy truck to result in a reported fatality is not likely to be the same as that of a single vehicle accident involving a smaller vehicle \( (Q_B) \). In Figure 1 we show how the proportion of heavy trucks among all vehicles in single vehicle fatal accidents changes as a function of the ratio \( Q_B/Q_A \). Recall that equation 6 has been derived assuming that, per unit of exposure, a medium/heavy truck as just as likely to be in an accident as any other vehicle. What the graph shows is that even if medium/heavy trucks have the same chance to be in a single-vehicle accident as any other vehicle, the proportion \( p \) of medium/heavy trucks amongst vehicles in reported accidents of a specified severity can be any number between 0
and 1. The magnitude of $p$ depends on the ratio $Q_A/Q_B$. Without an estimate of the ratio $Q_B/Q_A$, the knowledge of $p$ says nothing about the propensity of medium/heavy vehicles to be in accidents.

If medium/heavy trucks offer more protection against a fatality than lighter vehicles then $Q_B/Q_A > 1$. When $Q_B/Q_A > 1$ and the chance of a heavy/medium vehicle to be in an accident is the same as that of lighter vehicles, we should find $p < P$. That $p < P$ instead of $p = P$ as the equal-chance-assumption dictates, is due to fact that the two kinds of vehicles do not offer the same protection against fatal injury.

Consider next two-vehicle accidents. In this example we assume that $Q_{AA} = Q_{BB}$ and that $Q_{AB} = 3Q_{AA}$. That is, that while the chance of an accident between two medium/heavy trucks being reported as fatal is the same as between two smaller vehicles, the chance of an accident between a medium/heavy truck and a smaller vehicle is three times as likely to be fatal and so reported, as that between two vehicles of the same size class. Now we have $(1-P)/P = 0.93/0.07 = 13.29$ and therefore, by equation 7, $p = \frac{(1+13.29 \times 3)}{(1+2 \times 13.29 \times 3+13.29^2 \times 1)} = 0.19$. Thus, when the proportion of medium/heavy truck in fatal accidents is computed, (assuming that $Q_{AA} = Q_{BB}$ and $Q_{AB} = 3Q_{AA}$), it will be about 19%. This may look like gross ‘over-representation’ since the share of medium/heavy truck
in exposure was assumed to be only 7%. The difference between 7% and 19% is, in this example, wholly attributable to the fact that accidents between vehicles of unequal mass are more likely to be fatal than accidents between vehicles of similar mass. It is not an indication that medium/heavy trucks are more likely to be in accidents than lighter vehicles. What here appears as over-representation is a reflection of the difference in vehicle mass that is captured by the ratio \( \frac{Q_{AB}}{Q_{AA}} \).

One often encounters statements about entities of some kind being ‘over-involved’ in accidents. That is, that they are found in more accidents than is expected in light of their exposure. We have shown here is that over-representation and under-representation can occur even when the entities under scrutiny have the same probability of accident per unit exposure as any other entity. This comes about because accident frequency, the chance of an accident being reported, and accident severity, are all tied together in the measure of over-representation ‘\( p' \). Without knowing the Q-ratios \( \frac{Q_A}{Q_B}, \frac{Q_{AB}}{Q_{AA}}, \ldots \) one cannot know whether some noted over-representation is due to an increased probability to be involved in an accident per unit of exposure, or due to a difference in accident reporting, whether it is due to difference in accident severity, or perhaps any mixture of these three factors. This elementary ‘why-indeterminacy’ cannot be resolved except by obtaining independent estimates of the Q-ratios.

It matters whether over-representation is due to elevated probability of accident per unit exposure, due to larger likelihood of an accident to be reported, or due to a larger chance of an accident to be of a certain severity class. It matters to those who use over-representation as a good reason for intervention. The existence of the ‘why-indeterminacy’ counsels caution. Thus, e.g., if older drivers are over-involved mainly because their bones are brittle, because they are more likely to die as a result of trauma, or because they are more inclined to report accidents, one should not seek
reasons for their over-representation in declining cognitive skill, vision deficiencies or bad driving habits. In the same vein, should heavy trucks be found over-involved in severe accidents mainly due to their mass, one should not begin by seeking reasons or remedies in their stability, their brakes or the condition of truck drivers. Similarly, one may not attribute difference in the safety of roads solely to their geometric features if the roads also differ in the speed of travel (since speed affects severity).

Some argue (as one of the referees did) that no matter why an entity is over-involved, the fact that is over-involved is reason enough to initiate a search for remedies. By this argument, even if trucks are over-involved due to their mass and older drivers due to the brittleness of their bones, it is still sensible to attempt to “enhance the crash avoidance capabilities of trucks . . . and improve the cognitive skills of truck and older drivers. . .” . In my opinion, whether an attempt to enhance safety is or is not sensible depends largely on the expected cost of the attempt and on its expected payoff. The mere existence of over-representation without the knowledge of its cause can tell us nothing about either expected cost or its payoff. There is no logical reason to expect nor empirical evidence to suggest that indications of over-representation are associated with higher than average cost-effectiveness for measures aimed at reducing the chance of involvement in an accident.

In a research note C.D. Kemp (4) speaks of “An elementary ambiguity in accident theory”. It deals with a similar subject matter - partial ascertainment. Partial ascertainment arises when data censoring occurs so that not all events that occur are recorded. Kemp asks whether from data about recorded events it is possible say something about (the distribution) of events occurring. He concludes one cannot used recorded counts to separate the rate at which events are occurring from the process that censors it. In our case we have not only partial ascertainment but also differences
in severity that influence the severity category into which a reported event is placed. This results in
the “why-indeterminacy”. One cannot say whether over-representation is due to elevated frequency
of accidents, an above-average inclination to report an accident, or a higher severity-given-accident.
We argue that only when one can claim that all Q ratios are equal is it possible to give useful
interpretations to p. In all other cases one has to know the Q-ratios before p can be usefully
interpreted.

The message is not original. Evans (5) showed that when the susceptibility to die as a result of an accident is taken into account, the fatality rate for older drivers are dramatically lowered
and become similar to the rate for middle age drivers. Hakamies-Blomquist (6) refers to the same
issue as a problem of sampling in that: “... accidents of different driver groups may not be similar
samples of their total accidents, if the probability of an accident’s being reported varies among the
groups, or the inclusion criteria are met more frequently by some group’s accidents than by others.”

SUMMARY

Those who enter the thicket of road safety literature may find unexpected difficulties of
interpretation. Studies with seemingly correct methods at times produce contradictory findings. To
help these wanderers in the safety wilderness, this paper tries to shed light on two questions. The
first question is which of the many rates that have been proposed for measuring the safety of certain
vehicle types of driver groups is to be trusted. We conclude that apples and oranges should not be
mixed. For a rate to be correct, the numerator and the denominator must pertain to the same entity.
If vehicle exposure is in the denominator, then the count of vehicles of some type in accidents (and
not of accidents involving vehicles of some type) must be in the numerator; if driver exposure is in
the denominator then the count of drivers of some kind in accidents (not of accidents involving that
kind of driver) must be in the numerator. The second question is what meaning can be attributed to a finding of over-representation. If some type of entity is found over-represented in reported accidents of specified severity, may one conclude that entities of this type are in fact more likely be involved in accidents? Here the answer is ‘no’, and its root is in the distinction between accidents and reported-accidents-of-specified-severity. If one deals with reported accidents of specified severity (as is always the case), then over-representation may be caused by a mix of three factors: the probability to be in an accident per unit of exposure, the probability of the accidents to be reported, and the probability of the accident to be of the specified severity. Only if differences in the probability of an accident to be reported and the probability of an accident to be of the specified severity is accounted for (by the Q-ratios) can one get at over-representation due to differences in the likelihood to be in accidents. It follows, that an indication of over-representation cannot be taken to mean that the entity is over-represented in accidents and that one should seek remedies that reduce that chance to be involved in accidents.

REFERENCES


Figure and Table Captions:

**FIGURE 1** Fatal single-vehicle accidents

**TABLE 1** Counts of Accidents and Vehicles