THE DETECTION OF ACCIDENT-PRONE DRIVERS

By Dr H M Johnson

Highway Research Board

SYNOPSIS

Study has been made of the accident histories of 29,531 Connecticut drivers, selected at random, each of whom has been licensed in that State for a period of six years It was found that reported accidents were not distributed among these drivers according to the laws of chance, and that the discrepancy between fact and chance-expectation cannot be attributed to chance

The drivers whose records were examined had an average of 1 accident among 25 drivers in one year, which does not mean the same as one accident per operator in 25 years The contribution of accident-repeaters to the total number of accidents was quite large, and far in excess of what the laws of chance permit In fact, a group of less than 4 per cent of the operators had 40 per cent of the fatal accidents, 36 per cent of the non-fatal personal-injury accidents, and 38 per cent of the accidents which involved no personal injuries

By dividing the accident experience into two periods, it was found that operators who were accident-free during either period had the lowest rate during the other, those who had one accident during one period had twice as many accidents during the other period as those who went accident-free during the first Those who had four accidents each during one period had between 7 and 9 times as many accidents during the other period as those who went accident-free during the first period

Drivers younger than 21 years had a disproportionate share of all classes of accidents Their fatal-accident rate was nearly twice the average, and their non-fatal rate about 1½ times the average A census of the whole driver-population of the State showed, with respect to fatal accidents, the same disproportions throughout each of five years, the critical age was about 21 If these relations are nation-wide we could have saved about 3,100 lives in 1937 by bringing the fatal-accident rate of persons under 21 to the average rate of their elders About 7,800 lives could have been saved by bringing down the rate of drivers under 25 to the average rate of those over 25

Thus two classes of drivers have been found who are now accident prone those who are now young, and those who have had a high rate in the past

Consider a given population of automobile drivers From it you are asked to select a class that will include the largest possible number of those who will have more than a prescribed number of accidents within a specified period, and the smallest possible number of those who will not You will call these selected drivers "accident-prone" They will then undergo some special treatment They may be discharged from employment, or transferred to non-driving duties, they may be lectured to, psychoanalyzed, sympathized with, they may be reproved, exhorted, rebuked, they may be forbidden to drive upon the public highways, they may be subjected to inconvenient restrictions, they may be refused insurance or be required to pay an excessively high premium It is therefore important whom you select

How shall you proceed? Two methods are now under consideration The first takes account of the drivers' present behavior, the second of their past Let us call them the examination-method and the biographical-method, respectively. We do not have to choose between them, according to their merits, we may adopt either, neither, or both

The examination-method takes two forms. direct and indirect The direct form requires you to take each person out to a standard course, which presents the same hazards to all Let the hazards correspond to those of actual

traffic as nearly as possible Note how he deals with each of them and rate his performance accordingly Assume that those who are most inept in the test are also those who are most likely to cause trouble on the road This diagnosis presupposes, first that each person drives today in the test as he would drive today on the road, second that he drives on the road today as he will habitually drive during the period within which his accident-record is to be predicted To state these presuppositions is enough to cause them to be doubted The diagnosis depends on them

Consider now the examination-method in its indirect form Instead of having the driver operate a car under observation and under standard conditions, you subject him to a set of tests of particular skills, each of which is presupposed to be "necessary to safe driving", you subject him also to an interrogation about his "attitudes" or intentions Then, by combining his scores in the several tests in some manner which the testers have agreed upon, and by comparing the rating of each subject with that of every other, you formulate some rule, according to which you undertake to select those drivers who habitually drive the worst This procedure employs the same presuppositions as the direct method uses, and more besides If the rule of selection is valid, then there must be a usefully close association between the scores on the test and the accident-rates of the subjects Thus far, I have found no published report of a comparison which warrants the conclusion that test-performance is thus closely associated with accident-rate

In particular, the published reports of J -M Lahy, W Stern, Slocombe and Brakeman, Bingham, Sachs, and Tramm contain material which contradicts such a conclusion, even though some of these authors asserted it, while the published reports of Munsterberg, Miles and Vincent, C S Myers, and H M Vernon do not give the kind of evidence that is necessary to evaluate their conclusions that this is the case

In the United States, A R Lauer, at Iowa State College, and H R DeSilva and T W Forbes, at the Harvard Bureau for Street Traffic Research, have been developing different sets of tests which it is hoped will be related to accident rate Their reports published before this article was written do not enable one to conclude that they have yet succeeded, nor to conclude that they never will Our hopes, of course, are for their final success But in order to appraise the validity of any test or "battery" of tests, one must be able to compare the test-scores with some dependable criterion of driving performance which is independent of the test The best criterion, perhaps, is the operator's driving history—the duration of his experience, the amount and kind of his driving or his rate of encountering hazards, his accident-rate, etc To enable one to judge the tests requires records on some thousands of drivers; to collect and assemble the data ready for the statisticians is expensive, in one recent investigational series of tests about $6 50 per driver, the money has to be raised, and thus far there has been no rush of would-be contributors Hence, these investigators require time

Some attempts toward acceleration of these pre-appraisal activities have been made With the cordial and able cooperation of these investigators with the Highway Research Board, the U S Bureau of Public Roads, and the Commissioner of Motor Vehicles of the State of Connecticut, some 3,600 drivers were recently subjected to the Iowa State and the Harvard Bureau tests, in whole or in part The test-scores are now being correlated with one another, and also

with the annual number of reported accidents, determined from the official records of the Department of Motor Vehicles This study cannot be reported until it is completed and the remaining work will require several months Meanwhile, the most that should be said is that the appraisal of the tests is no simple matter

Strong claims have been made for the validity of certain sets of aptitude-tests designed for taxicab drivers—some by A J Snow, and others by David Wechsler, and for street-car drivers by Morris Viteles Inquiries addressed to the presidents of the employing companies brought out that Wechsler's tests have been abandoned by the company which first sponsored them, that Viteles' tests are still in use and highly regarded by the personnel department, on grounds which the correspondent did not analyze The inquiry about Snow's tests has not been answered On the whole, one should strive to be open-minded and skeptical about the value of all indirect tests, being predisposed neither toward premature rejection nor toward precipitate acceptance of the claims that have been made for their merits

Consider now the biographical method Instead of trying to ascertain how each individual drives on a given date, and thence to infer how he habitually has driven, now drives, or will drive, one may try to find out how he habitually drove in the past This presupposes a reliable accident-history

Some psychologists assume that one can get this information from the driver, by asking him questions, on one date, about a long history They presuppose that he has good intentions, but also a good memory—that his recall of past events will not be affected, for example, by an inconvenient tendency to forget all that might upset his belief that he is an excellent driver But in many drivers this belief is practically a systematic delusion, on this subject they are paranoic A much better way is to choose a population whose accidents have to be reported in detail, very soon after they occur, the testimony of more than one person, and also of the objective circumstances, being recorded Some large fleet-operators record these data, few, apparently, have made good use of them Among those who record and use them are certain utility companies and certain departments of the United States Government

These records, if properly kept, serve to identify a class of drivers who have repeatedly had accidents in the past, and the operators who are thus classified tend very strongly to give predictable histories, not so much as individuals as classified groups

Can such records as a state obtains enable one to select a class of drivers whose group-performance can be reliably predicted? The question is important To answer it, we chose the state which kept its records in the form most convenient for our purposes namely, Connecticut From a list of about 408,000 drivers known to have been licensed in 1931, our workers selected every 10th name This gave a "random" selection of 40,800 names From this selected list all drivers were eliminated unless the records of the department showed that they were licensed in every year in the period 1931-1936 This left a population of 29,531 drivers, for whom the official reports were complete for these six years Among them there accrued 7,082 accidents, which involved 5,650 operators If the accidents had been distributed among the operators irrespective of their personal identities and histories, then the distribution could be described by Poisson's law of small chances, for which the sole parameter is the rate of accidents per operator, within limits of inaccuracy

which are defined by the parameter itself If, within those limits, the law is not satisfied, then the distribution of accidents per operator is not impartial, but is subject to some systematic influences which remain to be determined experimentally

Table 1 shows that the actual distribution of these accidents does not follow the laws of chance Note that there is an excess of accident-free operators, and also an excess of repeaters, with a corresponding deficiency of operators having a moderate accident-rate We analyzed nine other populations in the same manner, and found that eight of them showed this characteristic Table 2, based on the driver-population studied by Slocombe and Brakeman, brings out the discrepancy in a more striking fashion These authors included all accidents, regardless of damage, and thereby obtained a much higher rate of accidents per operator than obtained in our population

Figure 1 shows what kind of predictions can be made from these histories Here we classify the operators according to the number of accidents which they had in 1931-1933, and show their accident-rates for 1934-36 Bear in mind that three years is a short time in which to establish an accident-history if we consider, as in this case, only those accidents which occur, in the average, once in 25 years of an operator's experience And yet, those who had one accident each in the first three years have twice as many accidents per operator in the second three years, as those who went through the first three years accident-free The probability is only $(10)^{-24}$ that the difference between these two rates is due to chance And those who had more than one accident in the first period have correspondingly higher rates in the second period

Figure 1 shows the converse relationship also Here we may see that not only can we predict the future performance of these classified groups from knowledge of their past performance, we can also infer their performance in the remote past from knowledge of what they have done in the less remote past In fact, the second kind of prediction

TABLE 1

ACTUAL AND EXPECTED DISTRIBUTION OF ACCIDENTS, INCLUDING CASUALTIES AND PROPERTY DAMAGE EXCEEDING $25, REPORTED TO THE COMMISSIONER OF MOTOR VEHICLES OF CONNECTICUT, 1931-36, IN A LICENSED DRIVER SAMPLE SELECTED AT RANDOM

Accidents per operator during experience	Operators having these accidents			Accidents accruing to these operators		
	Actual number	Expected number	Difference	Actual number	Expected number	Difference
0	23,881	23,234	647	0	0	0
1	4,503	5,572	−1,069	4,503	5,572	−1,069
2	936	668	268	1,872	1,336	536
3	160	53	107	480	159	321
4	33			132		
5	14	4	47	70	15	212
6	3			18		
7	1			7		
Totals	29,531	29,531	0	7,082	7,082	0

Note The probability that the differences between the actual and expected distributions is due to chance = $1\ 6\ (10)^{-161}$, which is insignificant

TABLE 2

ACTUAL AND EXPECTED DISTRIBUTION OF ALL ACCIDENTS REPORTED TO EMPLOYER, REGARDLESS OF PERSONAL AND PROPERTY DAMAGE, OF OPERATORS OF ELECTRIC RAILWAY CARS AND MOTOR BUSES, EMPLOYED CONTINUOUSLY THROUGH THE ONE-YEAR EXPERIENCE (1927)

REPORTED BY SLOCOMBE AND BRAKEMAN

Accidents per operator during experience	Operators having these accidents			Accidents accruing to these operators		
	Actual number	Expected number	Difference	Actual number	Expected number	Difference
0	217	101	116	0	0	0
1	326	315	11	326	315	11
2	569	493	76	1,138	986	152
3	458	514	− 56	1,374	1,542	−168
4	258	402	−144	1,032	1,608	−576
5	145	252	−107	725	1,260	−535
6	99	131	− 32	594	786	−192
7	86	59	27	602	413	189
8	46	23	23	368	184	184
9	27			243		
10	24			240		
11	24			264		
12	7			84		
13	5	10	86	65	103	935
14	2			28		
15	2			30		
16	2			32		
17	2			34		
18	1			18		
Total	2,300	2,300	0	7,197	7,197	0

Note The probability that the differences are due to chance = $4\ (10)^{-217}$

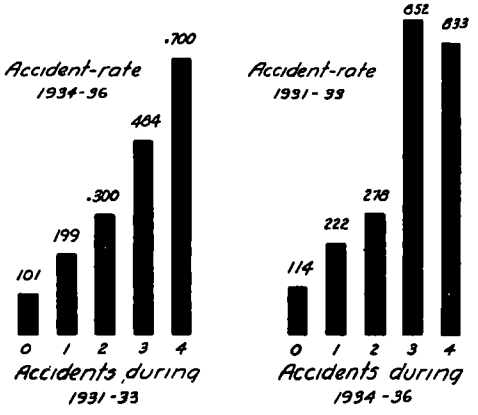

Figure 1 Relation of numbers of accidents in each half of the six year experience with the accident rates of the same groups in the other half

is more accurate than the first, perhaps because the younger drivers performed worse in their early experience, and so gave us more to predict

The accident-repeaters in this population were 3 88 per cent of the total Together they caused 39 per cent of the fatal accidents, 36 per cent of the non-fatal personal accidents, and 38 per cent of the non-personal accidents Thus, they have nearly the same proportion of every kind of accident For this reason the employer, the insurance company, and the licensing authority ought, if possible, to have a record of all accidents whether their consequences are important or trivial, in order to establish a reliable basis of prediction as early as possible

The accident-repeaters tend to shorten

the time between accidents as their accidents accumulate the fourth accident, for example, tends to follow the third more closely than the third follows the second

RELATION OF AGE TO ACCIDENT RATE

In this census the operators were classified according to their ages, which yielded some important information which is not evident in former compilations

Figure 2 shows that the drivers who were under 16-20 years old at the beginning of the experience and under 22-27 years old at its close had 1 47 times as many of the non-personal accidents as they would have had if the distribution of accidents were independent of age That this difference is not accidental is evidenced by the fact that the probability of the independency-hypothesis being true is less than $(10)^{-24}$ Figure 2 shows likewise that this same group had 1 53 times as many non-fatal personal injury accidents as the independency-hypothesis allows Moreover, this age-group had 1 83 times as many fatal accidents in this period as the independency-hypothesis allows Here the probability is as great as 1 in 250 that the hypothesis is true But, it is still a very conservative bet that the driver's accident-rate is influenced by his age That the odds are easier in the case of fatalities than in the other classes of accidents is due to the smaller number of individuals involved in fatal accidents

Thus, it looks as if those accidents which involve the youngest class of drivers are likely to be more severe than those which involve the older drivers

These findings led to a request for an age-census of all licensed drivers and all identified drivers involved in fatal accidents in Connecticut through the period 1932-36, that being the only period for which the individual accident-records had not been destroyed No state had made and published such a census up to that time The results are extremely valuable Figure 3 shows the age-distribution of the licensed drivers of Connecticut for 1929, and for each of the years 1932-36 Plainly, the average age is increasing, the proportion of youths becomes smaller in every year The probability that the populations of 1932

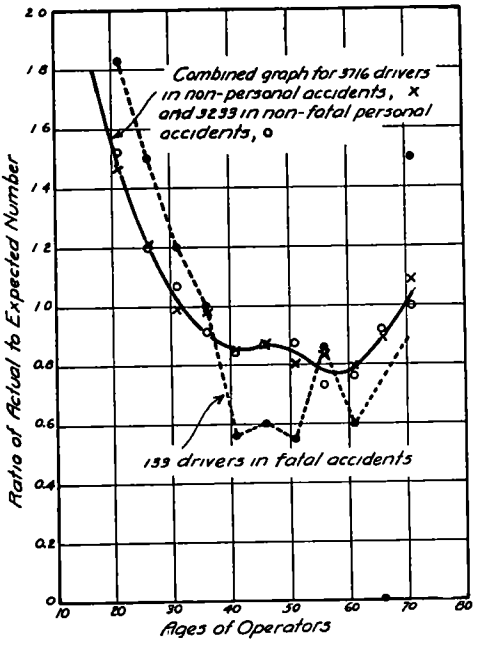

Figure 2 Relation of age to traffic accidents from histories of 29,531 operators selected at random from those licensed in Connecticut in every year from 1932 through 1936

and 1933 are homogeneous is only 1 in 250 in the next closest instance the probability is $(10)^{-24}$

Figure 4 shows the ages of 2,467 identified drivers who were involved in fatal accidents in Connecticut during the period 1932-1936 In this figure one year was chosen as the range of each age-class, in order to ascertain whether a critical age exists, especially in the youngest group Whether the data will bring out such a fact or not depends partly on the

number of persons in each of the neighboring age-classes This condition is satisfied for the earliest ages, and it is manifest from the graph that a critical

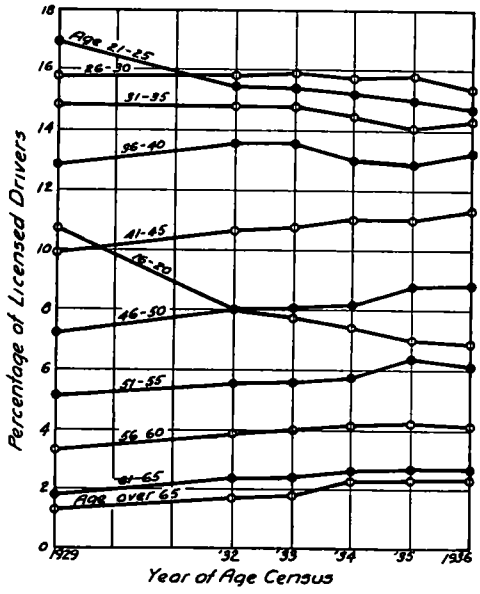

Figure 3 Relative number of drivers of various ages in Connecticut

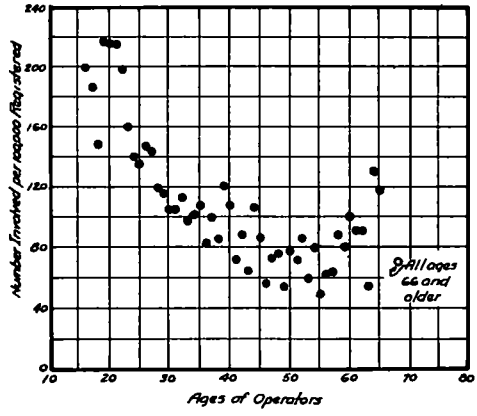

Figure 4 Drivers involved in fatal accidents per 100,000 drivers registered, Conn 1932–1936

period lies somewhere between the ages 19 and 21 Certain psychiatrists have suggested that the female population ought to show similar crises in the late 20's and the late 30's We have considered the two sexes separately, but since of the 2,467 identified drivers who were involved in fatal accidnts, only 136 were women—a little more than one-half of one per cent—and since the number of all drivers is by no means large for the purpose, it is not easy to make out any age-trends among the women, if such trends exist

Because the population is still rather small, we have grouped its members into 5-year age-classes, and have expressed the actual number of fatal accidents in each of these classes as a ratio to the corresponding number to be expected on the hypothesis that the rate per operator in each age-group is the rate for the population as a whole The black circles in Figure 5, and the curve fitted to them, indicate the result Thus, the age-group 16-20 years contains 1 73 times as many fatal-accident drivers as the hypothesis implies, the 21-25 year group contains 1 48 times as many, and so through The groups which contain the smallest proportion of involved drivers are 46-50 and 51-55 years The probability that these discrepancies between fact and hypothesis are jointly due to chance is $4(10)^{-48}$, which indicates practical certainty that the differences are statistically reliable

Disproportions of this same order appeared in every year of the experience taken singly, they also appear in three experiences from Massachusetts —in which, however, the basis of expectation is less reliable than in the present instance

The open circles in Figure 5 show the ratios between the actual and expected numbers of drivers suspended for speeding in the District of Columbia in the first five months of 1936, according to their ages This does not prove that these youths who have fatal accidents are the same youths who drive at excessively high speeds, but it does sug-

gest that the manner in which these youths drive may deserve looking into

In 1936 the youngest age-group had 2 10 times as many fatal accidents per 1,000 drivers as the whole population had, and 2 26 times as many per 1,000 drivers as their elders had This ratio is the highest that appeared in any of these years But there is no evidence that their rate is increasing, or diminishing, except according to chance The rates for the older drivers, however, declined in every year of the experience, in the average by about four per cent of the rate for the preceding year Thus, the youngest drivers are growing relatively worse, because their elders are driving better or driving less

If the age-relations among the drivers in Connecticut hold on a nation-wide scale, it would have been possible to save the lives of 3,085 persons who were killed on the highways in the United States in 1936, if by any means whatever, we could have brought the fatal-accident rate of the drivers under 21 years of age to the rate of their elders, likewise, it would have been possible to save about 7,787 of the 37,800 who were killed, if we had reduced the rate of the drivers under 25 to that of their elders

These findings point to the weakest of all spots in safety-education and in administration According to all tests of separate skills which are believed to be used in handling a motor vehicle, the highest average scores usually belong to the age-groups in which we find the highest rates of fatal accidents and also of personal injury accidents Moreover, it seems to be true that most drivers 20-22 years of age are able to handle a car on the road more skillfully than their elders It therefore seems that the question is not so much how skillfully a person can drive, as whether he will use the skills that he has Some skillful drivers, relying on their agility and alertness, may enter hazardous situations that are a little beyond their ability, whereas less skillful drivers, being aware of their weaknesses, may stay out of them The fact that the youths are the ones who have been most often and most intensively preached to in recent years ought

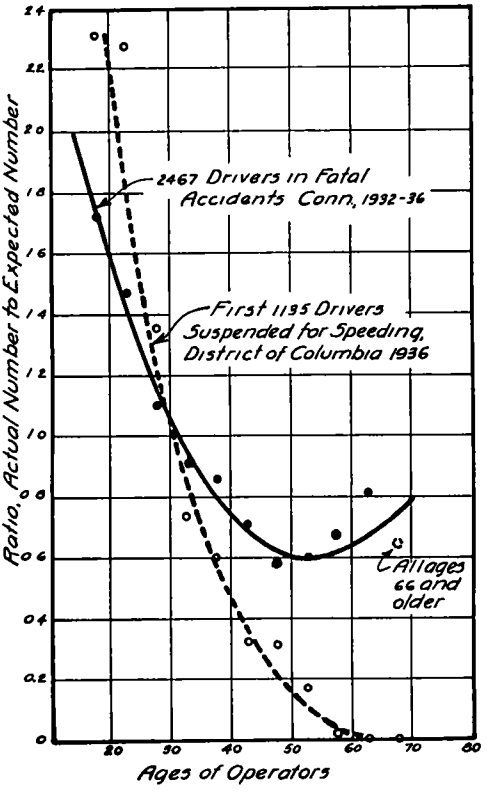

Figure 5 Comparison of ages of motor-vehicle drivers suspended for speeding in the District of Columbia with ages of drivers involved in fatal accidents in Connecticut

to suggest that our efforts in safety-education may be improvable

To summarize our biographical investigation has identified two classes of drivers who are now accident-prone They are, namely, those who are now young, and those who have had many accidents in the past

DISCUSSION ON ACCIDENT-PRONE DRIVERS

DR A R LAUER, *Iowa State College* There are a few points in Dr Johnson's paper on which I think it might be well to elucidate further relative to the methods involved Dr Johnson has presented the two methods, indirect and direct The biographical method would be all right if we had unlimited time to determine who should and who should not have a license We have had thrust upon us this problem which one cannot solve as has been suggested

I have never gone on record as supporting any of these tests Having analyzed several sets of similar data before, I venture a guess that when our final analysis is made we will find no high validity coefficients This is to be expected from the conditions and the phenomena with which we are working Human nature is unstable and highway conditions are highly variable Yesterday we heard talks on studies made of the maintenance costs of highways in different parts of the country As you remember, those maintenance figures varied from $100 to $500 or $600 a unit on different types of highways which apparently were comparable This is to be expected For similar reasons validity coefficients will never be high as long as we lump accidents into a unitary criterion They will not be high even if we segregate our accidents according to defective manipulation, defective vision, or how ever you may care to analyze them One reason for this is that there is some evidence that most accidents are caused by more than one condition In any one accident you may have three or four directly significant factors to consider We might mention the test for glare and accidents allegedly caused by glare The fact that one was blinded by glare would not mean necessarily that it caused the accident, but it is simply a factor which may or may not operate One may meet a thousand animals on the highways at night at such times that oncoming headlights are less blinding One will never get high validity coefficients so long as numerous factors produce the same end

Another point is this I have for many years maintained that the best test of a driver is 50 years of driving This would be the ideal test The best test of highway surface wear would be to lay the highway down and see how long it will last In the meantime your motor vehicle departments and others are faced with the practical problem of trying to keep certain groups of drivers from having so many accidents We can't wait until each one kills a few people to control his driving habits

I had lunch today with a man who lost his son He said go ahead and do what you can to reduce accidents but it is too late for me as I have lost my son If it is necessary to wait until a driver kills several people we certainly are not using the most economical method of dealing with the problem. What would we do with these young drivers? We cannot eliminate them from the highway for the simple reason that it is unconstitutional A remedy which might have some merits as practicable and which we will leave with you is that of legislation to reduce the speeds of certain persons or restrict their driving as some states have done at the present time Let a man drive according to his limitations

Now I have no brief to hold for tests as given by myself or anyone else As a matter of fact we have been changing our equipment constantly and you will find our clinic quite different than it was last fall We eliminate as soon as we find the test has no practical value However, as long as we are giving examinations and drivers are being charged

for them at the rate of $5 or so per examination, as some states are doing, why not do the best job we can?

We should be interested in this problem from the standpoint of trying to protect persons from injury rather than to dictate to drivers whether they can or cannot drive The medical profession do not guarantee to cure any ailment, they merely examine a person, tell him what his physical characteristics are, and with their experience, may be able to tell, within reasonable limits, what his chances are of living for a certain number of years There is never certainty, it is a matter of probability The insurance companies have been building up a rather thriving business based upon prognostic examinations of this type In a similar way I fully believe a scientific examination will help drivers stay out of trouble

The final point concerns the matter of expense In the present instance we have been given a big problem and it is necessary to be rather careful in everything we do It is not exactly typical of the expenses necessarily incurred from regularly examining drivers in large groups We examined 2,500 persons in the State of Iowa at a cost of little more than 60 cents per driver A year before we examined 1,000 commercial drivers covering 4,000 miles for $1 50 apiece The cost depends somewhat upon the number of people you are examining and the economical use of the equipment One should try to be open-minded I should say that we should ask people who offer any test for sale or rent, regardless of what it is, to show that it does better and is more economical than what has been used before If it does better than what has been used before, it is justified, if not, there is no reason to change This is a simple and unbiased answer as to the value of any licensing procedure

MR C M JOHNSTON, *Bureau of Public Roads* I should like to ask Dr Johnson if both people in an accident were counted in tabulating the statistics

DR JOHNSON These were involved drivers taken without regard to the distribution of responsibility I think most administrative authorities agree that you cannot distribute responsibility It is very seldom a man is involved in an accident except by doing something or failing to do something that would have kept him out of it

MR C M JOHNSTON, *Bureau of Public Roads* What I am trying to get at is, if you add the one who is 16 to 25 involved in the accident—he might have run into someone in the 40-year old group?

DR JOHNSON Then you have two involved, each one being counted

MR C M JOHNSTON, *Bureau of Public Roads* If you can take that accident from the 40-year group and leave it in the 16-year old group, then how would the picture look?

DR JOHNSON I do not know, that would require going back of the records that are available You would have to go into the analysis of the particular accidents which would take an expert to do and the facts are not available The number of deaths is about 0 97 times the number of involved drivers There are some few accidents that involve more than one driver and a few accidents involve more than one death for a given driver, but the ratio is very close to 1 to 1,—one accident, to one identified driver

DR LAUER I don't beleve these data shaped up as well as they deserved You did not have mileage on any of them

DR JOHNSON Not on any of them In another study of course we are recording some data that you obtained from nearly 3,000 drivers, namely their own reports of miles traveled I think that you will get an interesting age distribution You have already got that for that matter in your own work, but I don't believe I trust the reliability of those reports In the first place, if a man does not keep a record of his gasoline or of his speedometer, he does not know how many miles he drives in a year In the second place, he answers questions under circumstances that may invite him to exaggerate or to suppress and in the third place, everybody knows that a driver's hazards do not depend simply on his mileage In the fourth place, a very large portion of his hazards are hazards of his own creation

DR LAUER There is of course no doubt that it is reasonable to suppose that a salesman 30 or 40 years old who travels daily travels more miles per year than a youngster

DR JOHNSON Increased exposure may have something to do with increased rate, but I doubt very much if you can plausibly assume that children under 21 have a higher exposure rate than the average 30-35 year old group

DR LAUER I was assuming they had less

DR JOHNSON Then the youngsters are relatively worse than they now appear to be

PROFESSOR MORRISON To me there is no mystery as to why young drivers have many accidents, and I doubt if it has much to do with their driving experience, in years or miles Even casual observation shows that when a car shoots ahead in traffic the chances are at least ten to one that the driver is a young person These young drivers are usually skillful but they do not have, or at least do not exercise, judgment Youth has always been impetuous, daring, and rather careless of consequences Also, it is usually young people, rather than their elders, who go out for a good time at night, visit roadhouses and have enough drinks to cause accidents

The general assumption is that in all traffic accidents it is the drivers who are to blame, but I believe that in the majority of pedestrian accidents the primary fault lies with the pedestrians In one city half the pedestrians killed in 1936 were either crossing streets against red signals or crossing between intersections, usually in the dark It takes a very careful and alert driver to avoid hitting an old man in dark clothes crossing in the middle of a block on a poorly lighted street at night, and hardly anyone can avoid hitting a child who suddenly darts out in front of him from between parked cars