Accidents on Rural Two-Lane Roads: Differences Between Seven States

JERRY C. N. NG AND EZRA HAUER

Data on accidents, road characteristics, and traffic for rural, two-lane roads in seven states have been assembled. It was found that, for the same amount of traffic, different states record widely discrepant numbers of accidents. The discrepancy does not disappear even when roads with the same lane width, shoulder type, and terrain are examined. It is concluded that (a) data from different states should not be pooled, (b) warrants and standards based on accidents should be tailored to each state, and (c) the cause of the noted differences should be investigated.

To examine how accident occurrence is affected by lane width, shoulder width, shoulder type, curvature, and other characteristics of the road and how it depends on the amount of traffic flow, a large amount of data must be used, and they must be subjected to sophisticated statistical analysis. To secure a sufficient amount of data covering a wide range of conditions, it is common practice to pool data from several states. In this paper, the issue of whether data from different states can be combined is examined.

THE DATA

The data base used was assembled by Zegeer et al. and by Hummer (*1,2*). The data have been thoroughly checked and documented. They include information about the road, roadside, accident history, and traffic volume for almost 5,000 miles of two-lane roads collected from seven American states (Alabama, Michigan, Montana, North Carolina, Utah, Washington, and West Virginia). These states were selected to secure consistency in both accident reporting and coding policies. These 5,000 miles consist of 1,944 road sections. A road section is a stretch of road that is homogeneous with respect to lane width, shoulder width and type, and so forth. The roadway sections are ½ to 10 miles long. Of the 1,944 road sections, 1,801 are located in rural areas.

The majority of the road sections have a 5-yr history (1980–1984) of police-recorded accidents. Only the total number of accidents per road section (by type and severity) is available. Estimates of the average daily traffic (ADT) are for the sum of flows in both directions.

DIFFERENCES IN "ACCIDENTS PER MILE-YEAR" BETWEEN STATES

In Table 1 the pooled data from all seven states are used to examine how the "average number of single-vehicle accidents per mile-year" varies with ADT on rural two-lane roads. The somewhat irregular ADT ranges were selected so that each range has approximately one-fifth of all accidents. These averages are plotted in Figure 1.

The smooth curve in Figure 1 is the best fit to the disaggregate data when the model

$$\text{accidents/mile-year} = b_0\,(\text{ADT})^{b_1} \qquad (1)$$

is used. It appears that the change in single-vehicle accidents is nonlinear; the increase is sharp initially but tapers off as ADT becomes larger. This relationship agrees with the findings of previous research work as summarized by Satterthwaite (*3*).

The pooled road sections were separated by state to check whether the same accidents-versus-ADT relationship holds in all seven states, and the results are plotted in Figure 2. Although the points show considerable scatter, it is clear that there are major differences between the states. Thus, for the same ADT, for example, there seem to be three to four times as many single-vehicle accidents in West Virginia (filled triangles) as in Alabama (filled squares).

Similar accidents-versus-ADT plots are provided for head-on and sideswipe (opposite and same direction) accidents for each state in Figures 3, 4, and 5. The results are similar to that in Figure 2.

It appears that for the same ADT, different states record a markedly different number of accidents per mile-year. Unless this discrepancy can be attributed to other independent variables (lane width, shoulder type, terrain, etc.), it would have to be concluded that the pooling of state data is not advisable. The danger of pooling is that the relationship in Figure 1 could be an artifact of the composition of the sample and not a reflection of a real regularity in the relationship between accidents and traffic flow. Similar confounding in other variables could invalidate the results of statistical modeling.

It is therefore mandatory to go a step further to establish whether the differences in Figures 2 through 5 can be explained

TABLE 1 TABULATION OF SINGLE-VEHICLE ACCIDENTS VERSUS ADT

ADT Range	Mean ADT	Total No. of Accid.	Total Mile-Years	Ave. No. of Accid./M-Y	No. of Road Sections
0 - 1600	858	3629	10138	0.3580	644
1601 - 3010	2300	3444	5170	0.6662	400
3011 - 4550	3762	3546	3622	0.9791	309
4551 - 7000	5557	3554	2472	1.4377	235
7000 - 30000	10432	3507	1830	1.9169	193

Transport Safety Studies Group, Department of Civil Engineering, University of Toronto, Toronto Ontario M5S 1A4, Canada.

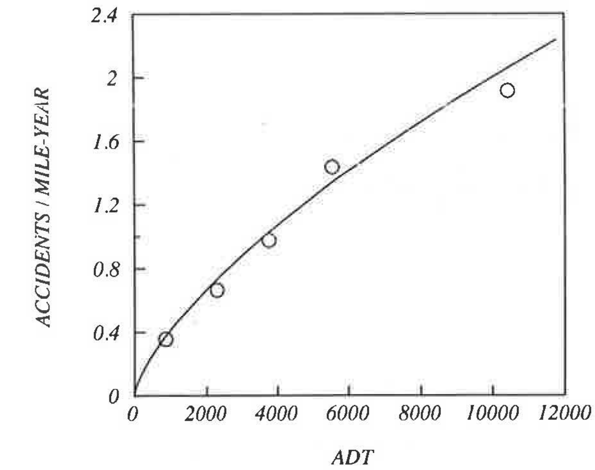

FIGURE 1 Accidents versus ADT, single-vehicle accidents, rural two-lane roads.

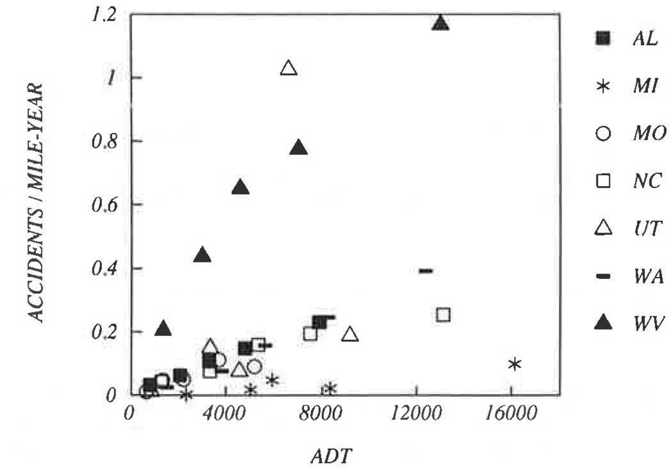

FIGURE 4 Accidents versus ADT, sideswipe (opposite direction) accidents, rural two-lane roads, by states.

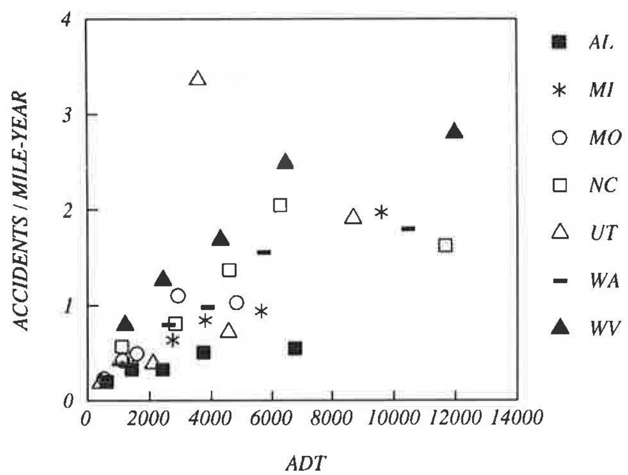

FIGURE 2 Accidents versus ADT, single-vehicle accidents, rural two-lane roads, by states.

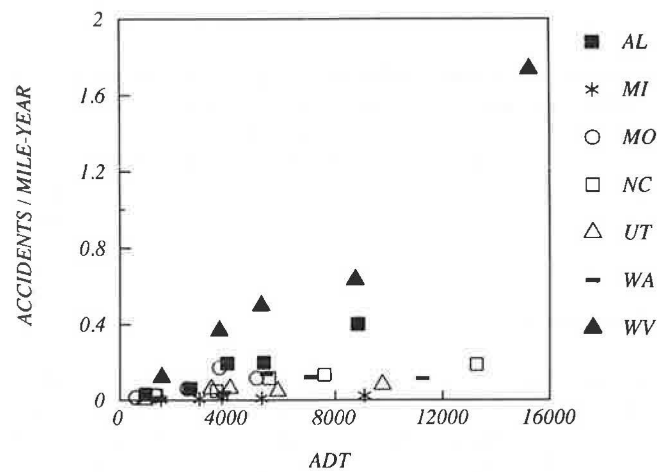

FIGURE 5 Accidents versus ADT, sideswipe (same direction) accidents, rural two-lane roads, by states.

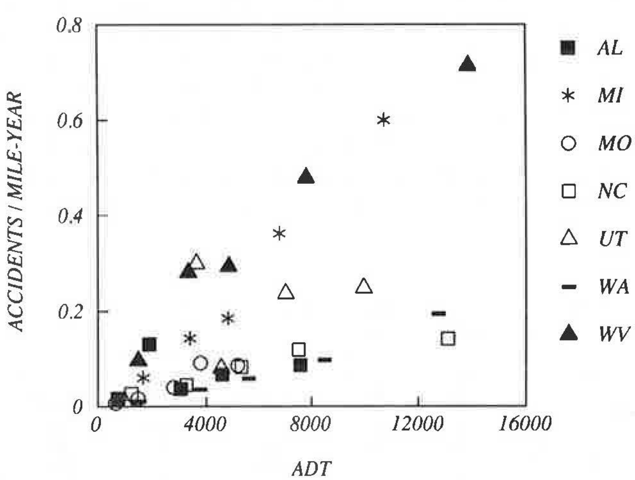

FIGURE 3 Accidents versus ADT, head-on accidents, rural two-lane roads, by states.

by the differences among states in lane width, shoulder width, terrain type, and so on.

ELIMINATION OF SOME INDEPENDENT VARIABLES

To determine whether the differences between Alabama and West Virginia are due to road conditions that we know about or to other, unknown factors, accidents-versus-ADT plots for road sections are compared with similar features. For a reliable comparison, an effort has been made to find that set of conditions that is found frequently enough in both states.

Only in the "rolling terrain" category are there enough road sections in both states. For this terrain, roads with 10- and 11-ft lanes are studied separately. Because the majority of the roads in both states do not have paved shoulders, only roads with unpaved shoulders are used. All road sections with a total unpaved shoulder width between 0 and 5 ft are used.

Thus, in Tables 2a and 2b, how the number of single-vehicle accidents varies with ADT for Alabama and West Virginia is examined using only rural two-lane roads in rolling terrain, with unpaved shoulders of 0 to 5 ft, and for lane widths of 10 and 11 ft. Because of the small samples, the road sections are grouped into four ADT ranges and the results plotted in Figures 6a and 6b. The bars in the diagrams are placed at two standard errors, corresponding statistically to a 95 percent confidence interval, above and below the estimated means. The smooth curves are the best fits to the disaggregated data when Equation 1 is used. It is clear from Figures 6a and 6b that West Virginia has consistently more single-vehicle accidents than Alabama for the same ADT, even after equalizing for road conditions.

Similar accident-versus-ADT tabulations and plots are provided for head-on and sideswipe accidents in Tables 3, 4, and 5 and in Figures 7, 8, and 9, respectively. In these, the lane width of 11 ft is studied because only for this condition are there enough road sections and accidents in both states for a

TABLE 2 TABULATION OF SINGLE-VEHICLE ACCIDENTS VERSUS ADT, RURAL, ROLLING TERRAIN, UNPAVED SHOULDER 0–5 FT

(a) LANE WIDTH = 10 FT.

ALABAMA			WEST VIRGINIA		
Mean ADT	Ave. No. of Accidents/Mile-Year	Std. Error	Mean ADT	Ave. No. of Accidents/Mile-Year	Std. Error
535	45/145.28 = 0.31	0.05	1467	137/144.65 = 0.95	0.08
964	36/122.16 = 0.29	0.05	2250	129/106.60 = 1.21	0.11
1636	52/ 81.57 = 0.64	0.09	4857	142/ 61.95 = 2.29	0.19
5002	33/ 31.98 = 1.03	0.18	11040	148/ 49.05 = 3.02	0.25

(b) LANE WIDTH = 11 FT.

ALABAMA			WEST VIRGINIA		
Mean ADT	Ave. No. of Accidents/Mile-Year	Std. Error	Mean ADT	Ave. No. of Accidents/Mile-Year	Std. Error
1338	59/177.12 = 0.33	0.04	1750	60/ 35.45 = 1.69	0.22
2101	65/121.21 = 0.54	0.07	3633	53/ 41.05 = 1.29	0.18
3394	56/ 95.75 = 0.58	0.08	4450	60/ 19.55 = 3.07	0.40
7046	64/ 84.21 = 0.76	0.10	8300	66/ 25.10 = 2.63	0.32

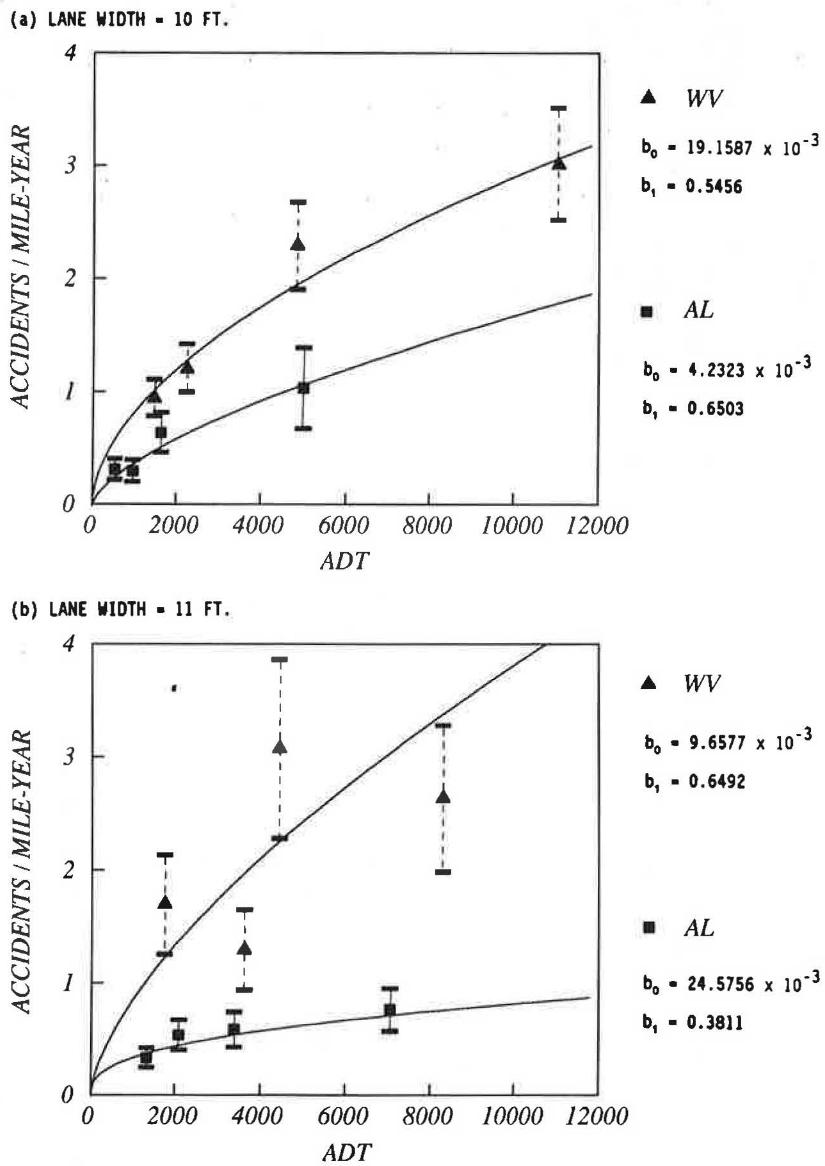

FIGURE 6 Alabama and West Virginia, single-vehicle accidents, rural two-lane roads, rolling terrain, unpaved shoulders 0–5 ft.

TABLE 3 TABULATION OF HEAD-ON ACCIDENTS VERSUS ADT, RURAL, ROLLING TERRAIN, UNPAVED SHOULDERS 0–5 FT, LANE WIDTH 11 FT

ALABAMA			WEST VIRGINIA		
Mean ADT	Ave. No. of Accidents/Mile-Year	Std. Error	Mean ADT	Ave. No. of Accidents/Mile-Year	Std. Error
1577	6/267.33 = 0.02	0.01	1750	7/ 35.45 = 0.20	0.07
3281	9/126.75 = 0.07	0.02	3633	5/ 41.05 = 0.12	0.05
4942	8/ 46.68 = 0.17	0.06	4633	9/ 29.02 = 0.31	0.10
8550	8/ 37.54 = 0.21	0.08	9400	6/ 15.45 = 0.39	0.16

TABLE 4 TABULATION OF OPPOSITE DIRECTION SIDESWIPE ACCIDENTS VERSUS ADT, RURAL, ROLLING TERRAIN, UNPAVED SHOULDERS 0–5 FT, LANE WIDTH 11 FT

ALABAMA			WEST VIRGINIA		
Mean ADT	Ave. No. of Accidents/Mile-Year	Std. Error	Mean ADT	Ave. No. of Accidents/Mile-Year	Std. Error
1482	15/216.93 = 0.07	0.01	1750	13/ 35.34 = 0.37	0.10
2938	18/159.46 = 0.11	0.03	3633	9/ 41.05 = 0.22	0.07
4687	20/ 60.67 = 0.33	0.07	4633	14/ 29.20 = 0.48	0.13
8121	13/ 41.23 = 0.32	0.09	9400	23/ 15.45 = 1.49	0.31

TABLE 5 TABULATION OF SAME DIRECTION SIDESWIPE ACCIDENTS VERSUS ADT, RURAL, ROLLING TERRAIN, UNPAVED SHOULDERS 0–5 FT, LANE WIDTH 11 FT

ALABAMA			WEST VIRGINIA		
Mean ADT	Ave. No. of Accidents/Mile-Year	Std. Error	Mean ADT	Ave. No. of Accidents/Mile-Year	Std. Error
1577	18/267.33 = 0.07	0.02	2060	12/ 46.40 = 0.26	0.07
2964	15/ 88.04 = 0.17	0.04	4125	16/ 49.65 = 0.32	0.08
4382	23/ 85.38 = 0.27	0.06	5200	8/ 12.90 = 0.62	0.22
8550	18/ 37.54 = 0.48	0.11	11400	24/ 12.20 = 1.97	0.40

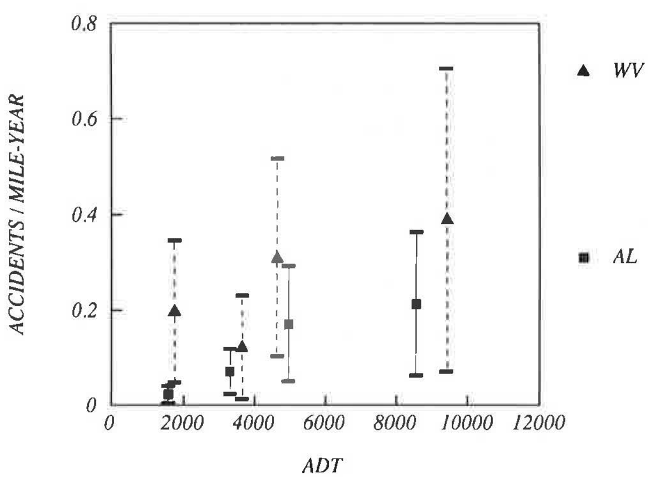

FIGURE 7 Alabama and West Virginia: Head-on accidents, rural two-lane roads, rolling terrain, unpaved shoulder 0–5 ft, lane width 11 ft.

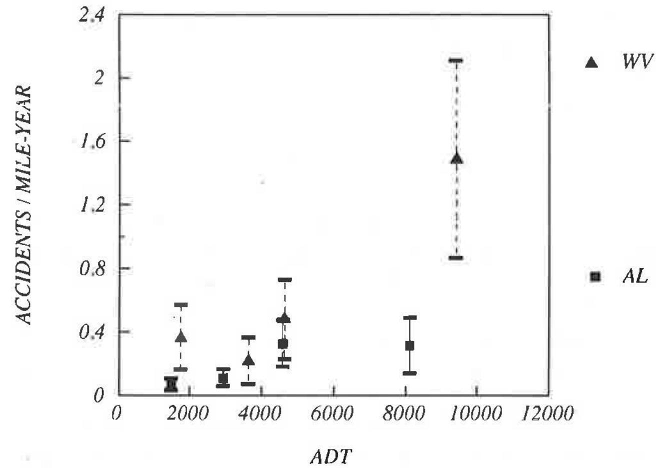

FIGURE 8 Alabama and West Virginia: Opposite sideswipe accidents, rural two-lane roads, rolling terrain, unpaved shoulder 0–5 ft, lane width 11 ft.

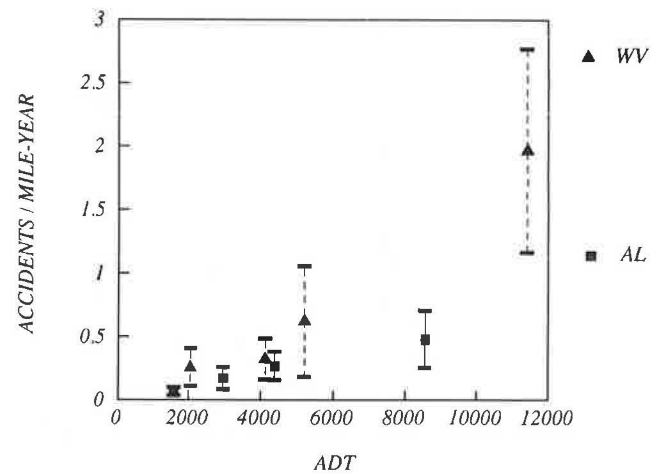

FIGURE 9 Alabama and West Virginia: Sideswipe accidents, rural two-lane roads, rolling terrain, unpaved shoulder 0–5 ft, lane width 11 ft.

reliable comparison. For these accidents no model has been fitted to the data. The relationship between traffic and multivehicle accidents will be a subject matter of a subsequent report. The results of Figures 7, 8 and 9 are as before, under similar conditions (ADT, terrain, lane width, shoulder width and type); West Virginia records more accidents per mile-year than does Alabama.

CONCLUSIONS

Under similar conditions different states have different average numbers of accidents per mile-year. Because we cannot account for these differences, and they are large, we conclude that data from different states should not be pooled for use in multivariate analyses.

More important, the existence of large, unexplained differences invites investigation. Are they a reflection of differences in accident reporting criteria or variable degrees of accident reporting (*4*), or do they contain hints to important differences in highway design or traffic management from which we could learn? It is the kingpin of epidemiology to identify the differences and search for their cause. The cause of the important differences noted should be found.

As a corollary, because of the large differences that are not currently explained, safety standards, warrants, and procedures that are based on accident frequency or rate should be tailored to each state individually. A nationwide accident warrant seems to make little sense when, under seemingly identical conditions, one state has on the average three to four times as many accidents as another.

ACKNOWLEDGMENTS

Work on this subject has been made possible by the support of the Federal Highway Administration, U.S. Department of Transportation. Their continued support and help are gratefully acknowledged.

REFERENCES

1. C. V. Zegeer, J. Hummer, D. Reinfurt, L. Herf, and W. Hunter. *Safety Effects of Cross-Section Design for Two-Lane Roads—Volume I—Final Report.* FHWA/RD-87/008. FHWA, U.S. Department of Transportation, 1987.
2. J. Hummer. *Safety Effects of Cross-Section Design for Two-Lane Roads—Data Base User's Guide.* FHWA, U.S. Department of Transportation, 1986.
3. S. P. Satterthwaite. *A Survey of Research into Relationships Between Traffic Accidents and Traffic Volumes.* TRRL SP 692. Transport and Road Research Laboratory, Crowthorne, England, 1981.
4. E. Hauer and A. S. Hakkert. *The Extent and Some Implications of Incomplete Accident Reporting.* Presented at the 66th Annual Meeting of the Transportation Research Board, Washington, D.C., January 1987.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.

DISCUSSION

CHARLES V. ZEGEER AND J. RICHARD STEWART
Highway Safety Research Center, University of North Carolina, CB #3430, Chapel Hill, N.C. 27599.

The paper by Ng and Hauer, "Accidents on Rural Two-Lane Roads: Differences Between Seven States," addresses an important issue relative to whether state data should ever be "pooled" for use in large-scale accident analyses. Accident experience by accident type is shown for all seven states (i.e., Alabama, Michigan, Montana, North Carolina, Utah, Washington, and West Virginia) before consideration of differing roadway features by state. Then the report focuses on a comparison of the single-vehicle accident experience between only two of the seven states, Alabama and West Virginia, for specific data subsets. The authors found unexplained differences and concluded that "data from different states should not be pooled for use in multivariate analyses." On the basis of our independent analysis of that same data base, however, we would offer some further analysis results and conclusions.

Before selecting these seven states for use in the initial research study (*1*), it was recognized that the 50 American states have varying degrees of accident reporting thresholds, reporting jurisdictions, accuracy for reporting the accident locations, and other accident data characteristics. For example, states that comply with a low accident reporting threshold (e.g., \$100 to \$200 per accident) may have considerably more property damage only (PDO) accidents reported (all else being equal) than states with a higher reporting threshold (e.g., those that report only towaway accidents and injury and fatal accidents). These and other factors were carefully considered when these seven states originally were selected for data collection. Accident and other data from the states determined the seven to be among the best states in terms of relatively good data quality and consistency. The large differences in geography, climate, terrain, and other factors among the seven states were recognized and considered desirable, so that the study results would represent a wide range of roadway, traffic, and other conditions found in the United States.

One of the first steps to test for differences between state data bases could be to explore overall average accident experiences (e.g., rates, severities, and types) in each of the seven states. Differences in accident reporting levels, as well as differing geometric and roadway conditions, could account for differences in overall accident statistics between states. The same basic data base was analyzed as that used by Ng and Hauer except for some minor adjustments made in the data base in recent years (e.g., 1,940 sections instead of 1,944 were used because of the omission of four high-volume sites with widely varying ADT throughout the section). The mean total accident rates in accidents/MVM ranged from 1.82 in Washington to 4.01 in West Virginia, as shown in Table 6. Overall average rates for the other five states ranged from 1.99 (Montana) to 2.82 (Michigan). Fatal accident rates ranged between .026 and .044 for five of the states, with higher values (.060 and .064) for West Virginia and Montana. The rate of injury accidents ranged from 0.66 to 1.00 for six of the states, with West Virginia again high at 1.63 (injury accidents per million vehicle miles). The rate of single-vehicle accidents was considerably lower for the Alabama sample sections (0.54) compared with the other states, highest in West Virginia (1.43), and relatively similar in the other five states (0.83 to 1.16).

To understand these accident trends better, it is useful to review the roadway and traffic characteristics of the data samples in the individual states. For purposes of this discussion, we have summarized average values for some key traffic and roadway features in Table 7. The averages of AADTs in the state samples range from 1,720 for Montana sites to 4,765 in North Carolina for these state-maintained roadway sections. West Virginia sites can be seen to have among the most restrictive geometrics, particularly in terms of narrow lanes (average of 10.4 ft), hazardous roadside conditions (4.6, where 7.0 is the most hazardous and 1.0 is the least hazardous), and the largest amount of sharp curves (i.e., 39.9 percent of the West Virginia sample has horizontal curvature of 2.5 degrees or greater). Thus, the combination of sharp curves, dangerous roadside, and narrow lanes would lead one to expect a higher experience of accidents than state samples with less severe roadway designs.

TABLE 6 SUMMARY OF ACCIDENT STATISTICS FOR SAMPLE SECTIONS IN SEVEN STATES

Accident Statistics	Alabama	Michigan	Montana	North Carolina	Utah	Washington	West Virginia
Sample Size (Number of Sections)	437	282	168	273	203	231	346
Rate of Single Vehicle Accidents (Acc/MVM)	0.54	0.83	1.16	0.95	1.11	0.86	1.43
Rate of Non-Run-Off-Road Accidents (Acc/MVM)	1.92	1.99	0.83	1.53	1.20	0.96	2.58
Rate of Fatal Accidents (Acc/MVM)	0.037	0.026	0.064	0.044	0.044	0.032	0.060
Rate of Injury Accidents (Acc/MVM)	0.66	0.78	0.86	1.00	0.78	0.80	1.63
Rate of Total Accidents	2.46	2.82	1.99	2.48	2.31	1.82	4.01

TABLE 7 SUMMARY OF ROADWAY CHARACTERISTICS FOR SAMPLE SECTIONS IN SEVEN STATES

Roadway Characteristics	Alabama	Michigan	Montana	North Carolina	Utah	Washington	West Virginia
Average Annual Daily Traffic (AADT)	2,978	3,182	1,720	4,765	2,380	3,713	4,619
Lane Width (feet)	10.5	11.3	11.5	10.6	12.3	11.1	10.4
Shoulder Width (feet)	5.9	8.2	1.9	6.6	3.0	5.8	4.1
Roadside Hazard Rating	3.7	3.6	3.5	4.1	3.8	4.1	4.6
Horizontal Curvature: Percent of section with 2.5 degree of curve or greater (percent)	10.9%	4.8%	8.8%	18.0%	25.0%	16.2%	39.9%

Ng and Hauer found differences in specific accident types between state data samples when accidents were plotted against ADT. However, this is not surprising, at least partly because of the differences in roadway conditions between the state samples. For example, as might be expected, the single-vehicle, head-on, and opposite direction sideswipe accidents were quite high in West Virginia compared with the other states. This may be expected as a result of the generally curvy, narrow roadways with more hazardous roadside conditions for the West Virginia sample compared with the other states. The incidence of opposite direction sideswipe accidents was quite low in Michigan, as might be expected; the Michigan sample sites have the widest combined width of lanes (11.3 ft) plus shoulders (8.2 ft), as well as only 4.8 percent of horizontal curves (which was the mildest horizontal curvature of the seven states). Thus, wide, relatively straight roadways would be expected to result in a relatively low incidence of opposite direction sideswipe accidents, as the data showed.

Although some of such variation in accident types can be explained by roadway and traffic differences, there are some differences in accident reporting, driver behavior, and so on that can also cause differences in accident types. One example is the low rate of reported single-vehicle accidents in Alabama. Although the total accident rate in the Alabama sample sections was right about in the middle of the seven states, the rate of single-vehicle accidents was the lowest of the seven states. One likely reason is the fact that accident types used for the seven-state data base had to be developed into a common definition based on the different accident report forms in each state. The Alabama accident report form does not have a specific code for run-off-road accidents, so a combination of several accident variables had to be used to classify each accident. By selecting data subsets and further dividing the data by specific accident type, state, roadway geometrics, and ADT category (as done by Ng and Hauer), there is also the likelihood of creating some cells with relatively small sample sizes of specific accident types where unreliable accident rates may result.

We would like to point out that on the basis of past research, accident relationships with roadway geometrics generally differ considerably between urban and rural areas. In fact, all of the computer modeling conducted in the initial research study by Zegeer et al. (*1*) was based on analysis of rural samples only. The analysis by Ng and Hauer apparently combined the high-volume urban sections with rural sections in developing their accident rate figures by state and ADT. Much of the spread in their accident rates for the seven states is in the highest ADT groups (i.e., above 10,000), where sample sections are mostly urban and where very small sample sizes exist (only 166 miles or 3.3 percent of the data base is urban).

In particular, for ADTs of 8,000 or less (which are nearly all rural roadways), accident levels for most accident types are much more consistent between the seven states.

It is also questioned why a measure of roadside hazard (i.e., either roadside hazard rating or roadside recovery area distance as contained in the data base for each section) was not used as a control variable by Ng and Hauer when comparing single-vehicle accidents between states. Roadside hazard rating (or recovery distance) was found in the original research study (*1*) to be the most important roadway factor (except for ADT) in explaining single-vehicle and other related accident types. By ignoring the roadside condition, unexplainable differences in single-vehicle accidents would surely occur when comparing state data samples that have differing levels of roadside hazard.

It should also be mentioned that the distribution of accidents by type can vary considerably depending on many roadway features. For example, if the data sample in State A has more intersections and driveways than the sample in State B, then one would expect a higher percent of right-angle, rear-end, and turning accidents for the data sample in State A than in State B. Such a condition would not necessarily require separating the data sets for analysis. Instead, one may use a measure of intersection or driveway frequency as part of the analysis.

Ng and Hauer conclude that because they found large differences between state data that they could not explain, they recommend setting standards, warrants, and procedures on the basis of accident frequency and rate to be "tailored to each state individually." This recommendation would appear to assume that accident data within a given state will be stable and consistent. Unfortunately, many differences exist within some states in terms of their reporting criteria (e.g., the city of Detroit investigates and reports a much smaller percentage of noninjury accidents than are reported in many other Michigan areas).

Further, some states have greatly differing terrain (mountainous areas and flatlands) and amounts of rain and snow, and even greatly differing driver characteristics (e.g., tourists versus mostly local drivers), depending on the area of the state. Thus, differences in accident experience may occur on roads in a state that may not be explainable by traffic and roadway variables alone. Such intrastate differences could cause equal or larger variations in accident experience than reported by Ng and Hauer between states. Does this mean that we must split each state's data into many data subsets before conducting accident analyses? We believe that it is reasonable to pool data for some states or jurisdictions but not for others on the basis of the characteristics of the data sets in question and the purposes of the analyses.

The point of Ng and Hauer that substantial differences may exist in data obtained from different states is well taken and, certainly, state differences should be investigated. When variables and relationships seem comparable across states, however, then analyses of combined data sets should yield estimates of relationships that are, in a sense, smoothed over a broader range of conditions and, hence, may be more widely applicable than those obtained from data within a single state. As Ng and Hauer pointed out, when major differences between states are found, they may suggest certain other factors that should be considered or potential problems with certain variables or data systems. Even when differences do exist, it still may be possible to smooth certain relationships across states while allowing others to differ from state to state.

For example, weighted log linear regression models were fit to subsets of the seven-state base to investigate relationships between ran-off-road accidents and roadway/roadside factors, such as ADT, lane width, shoulder width, recovery distance, roadside hazard rating, and terrain. Initial analyses indicated that the distributions of the relevant variables were similar and that it was reasonable to pool the data for the states of Michigan, Montana, North Carolina, Utah, and Washington. Alabama had much lower ran-off-road accident rates than did the five-state group, and West Virginia had much higher ran-off-road accident rates.

To investigate further the nature of these differences, the data from West Virginia were combined with the five-state data and analyses carried out; a significant state effect was found for West Virginia but no significant interactions. This suggests that although the magnitude of ran-off-road accidents is higher in West Virginia than in the other five states, relationships between accident rate and the roadway/roadside characteristics are similar. With Alabama data, on the other hand, significant interactions indicate that rates of single-vehicle accidents not only differ in magnitude from those of the other states but also that the nature of the relationships with the roadway/roadside factors is quite different. This could well be the result of problems in classifying ran-off-road accidents in Alabama, as discussed earlier, though not a particular problem in data for the other six states. One might reasonably conclude that pooling data from six of the seven sites may be quite reasonable for analysis of single-vehicle accidents. Further testing may well show that Alabama data may appropriately be combined with that of the other six states for analysis of total accidents and/or certain other accident types (particularly as the average total accident rate for Alabama was near the middle of the range of accident rates for the seven states).

In conclusion, we would again compliment Ng and Hauer on their addressing a very timely issue, that is, whether to combine data from several states. We do not agree, however, that state data bases should never be combined for analysis purposes. Instead, we believe that certain criteria should be used to determine whether two or more data sets should or should not be combined for a particular analysis. For example, such criteria may be expressed in the following questions.

1. Are the data variables defined consistently?
2. Are accident reporting thresholds reasonably similar? If not, it may still be reasonable to pool data from two states with differing reporting of property damage accidents and analyze only the injury and fatal accidents (if all other criteria are met).
3. How detailed does the analysis need to be? Do available data variables provide for sufficient accuracy for the intended analysis?
4. Is there a need to combine data from various geographic areas, regions, climates, and so on for a global accident analysis; or is analysis of a single state, city, or county sufficient?

We appreciate the opportunity to provide these comments and welcome other thoughts and further research on this timely subject.

AUTHORS' CLOSURE

In the first part of their discussion Zegeer and Stewart provide important further detail about the data that served as a basis of this paper and also of an earlier work (*1*). Eventually they conclude that the differences in accidents/mile-year that we show to exist are "due at least in part to the differences in roadway conditions between the state samples." They state, "Although some such variation in accident types can be explained by roadway and traffic differences, there are some differences in accident reporting, driver behavior, and so on that can also cause differences in accident types."

We of course concur with this conclusion. What we have tried to show is that even after one does account for differences in road conditions, large differences in accidents/mile-year still remain. That there are differences between the states is evident from Figures 2 to 5. That only a small part of the difference is due to "roadway and traffic differences" we show in Figures 6 to 9. In these we have compared road sections that are all in the same terrain, have the same lane width, same type of shoulder and shoulder width. The comparison is always between sections which serve the same ADT.

As the discussants note, roadside hazard and curvature have been accounted for only indirectly by comparing road sections that all are in a "rolling terrain." This is certainly a deficiency. We could not account for curvature directly because for more than half of the road sections in the data base this piece of information is missing. This is also why Zegeer et al. (*1*) did not use curvature as an independent variable. To examine further the effect of roadside hazard, we repeated the analysis for Figure 6, this time ensuring that the average road hazard rating and curvature for the Alabama and West Virginia road sections are very similar. The results are shown in Figure 10. Comparing it with Figure 6, the difference for 10-ft lanes is diminished but for 11-ft lanes it remains virtually as before.

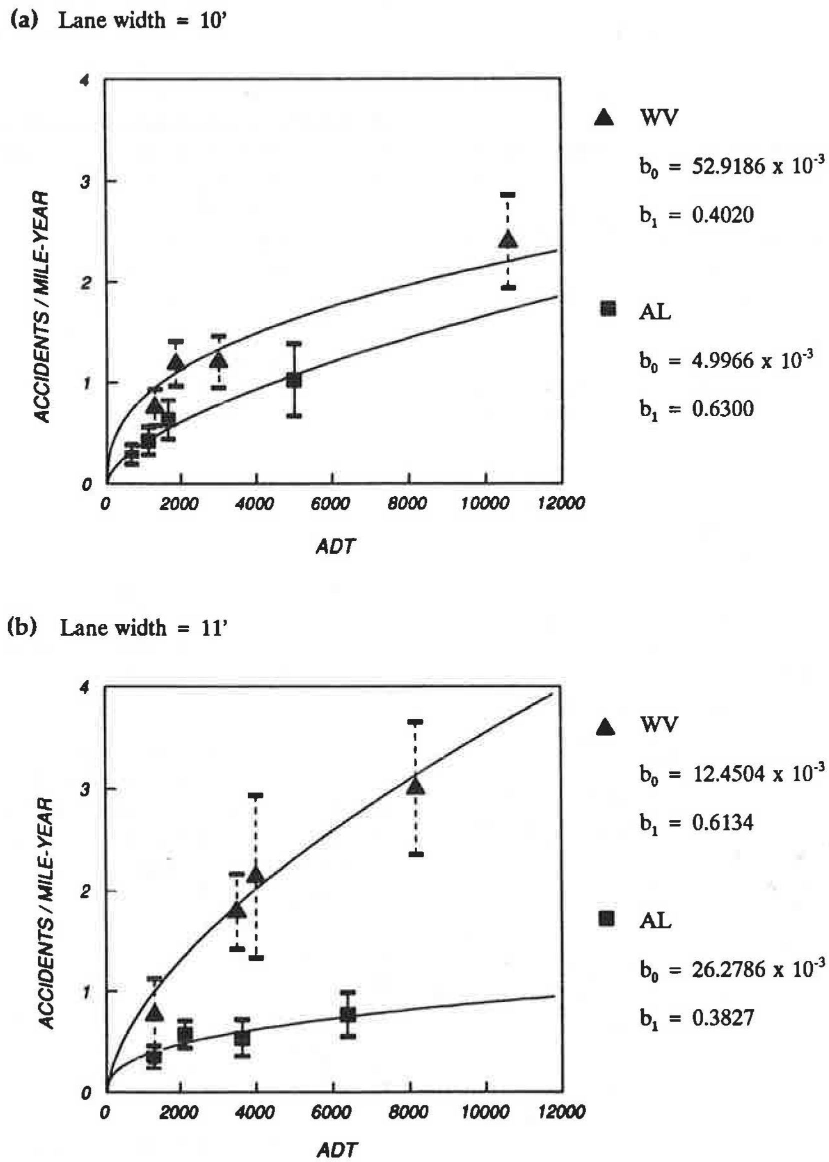

FIGURE 10 Alabama and West Virginia: Single-vehicle accidents, rural two-lane roads, rolling terrain, unpaved shoulders 0–5 ft.

Thus, the difference between the two states remains large, and neither the road geometrics nor the traffic flow about which we have data suffice to explain the differences in accidents/mile-year. Zegeer and Stewart seem to come to the same conclusions when they eventually say that in a multivariate model "Alabama had much lower ran-off-road accident rates than did the five-state group, and West Virginia had much higher ran-off-road accident rates." In addition, they find that "a significant state effect was found for West Virginia" and that for Alabama, "rates of single-vehicle accidents not only differ in magnitude from those of other states, but also that the nature of the relationship with the roadway/roadside factors is quite different." It appears that, reluctantly, the discussants agree with the observation that different states seem to have a different number of accidents per mile-year even when the road and traffic conditions appear to be similar.

We state in the paper: "Under similar conditions different states have different average numbers of accidents per mile-year. Because we cannot account for these differences, and they are large, we conclude that data from different states should not be pooled for use in multivariate analyses." Zegeer and Stewart seem to take issue with this conclusion when they say: "We believe that it is reasonable to pool data for some states or jurisdictions but not for others on the basis of the characteristics of the data sets in question and the purposes of the analyses."

Of course, it always true that data can be pooled for "some states or jurisdictions." To be specific, data can be pooled for those states and jurisdictions to which the same multivariate model can be fitted. However, when "there is a significant state effect" or "when the nature of the relationship with roadway/roadside factors is quite different," to pool data is perilous.

To illustrate, consider State *X* in which roads have an average roadside hazard rating of 3.7 and 40 percent of reportable accidents are reported, whereas in State *Y* the average roadside hazard rating is 4.6 and 80 percent of the reportable accidents are reported. In all other respects the roads and traffic in *X* and *Y* are very similar. The difference in the extent of accident reporting alone will cause State *Y* to have twice as many reported accidents/mile-year as State *X*. However, in a multivariate analysis in which data for *X* and *Y* are simply pooled, this will be seen as caused by the difference of 0.9 in the average roadside hazard rating. Thus the importance of roadside hazard will be exaggerated and money may be misspent.

The West Virginia road sections in the data set used here and elsewhere (*1*) have an average roadside hazard rating of 4.6 but for Alabama it is 3.6. (see Table 7). At the same time a West Virginia road section will have up to four times as many accidents as an Alabama road section with the same traffic and geometrics (see Figure 6 and Figure 10). If now data for West Virginia and Alabama are pooled, is there not a danger that in the ensuing multivariate model the real and the fictitious are inextricably mixed?

In no way do we intend to imply that the results of Zegeer et al. (*1*) are incorrect. This is impossible to say without a reanalysis of the data, which is now in progress. Our intent was only to point to the large differences between the seven states and to show that information about traffic and geometrics is not sufficient to explain it.