Pedestrian Exposure to Risk in Housing Areas

D.H. CROMPTON

ABSTRACT

A large-scale study has been carried out for the U.K. Transport and Road Research Laboratory to determine levels of pedestrian activity in representative housing areas and to examine their influence, along with other factors, on annual pedestrian casualty rates. Data were collected on land use and layout, population and socioeconomic characteristics, number of pedestrians (by age and sex), traffic, and casualties to pedestrians in 474 squares of 1 km each, distributed in the regions of England and Wales. Analysis of these data has resulted in a group of models in which annual casualty rates per square kilometer of housing area are explained in terms of pedestrian and traffic data, population and census data, and land use and layout data. The best of the models (which were tested against an independent data set) explain up to 77 percent of the variation about the mean casualty rate (R = 0.88). But standard errors of the estimate are disappointingly high. Factors influencing the size of these errors are examined, and possible practical applications of the models are discussed.

Annual records of accidents or casualties involving pedestrians in housing areas can be expressed as rates per 100,000 residents and used to compare conditions in different localities. These casualty rates vary considerably from region to region. Among the many factors influencing these variations, levels of pedestrian exposure appear likely to be important. The level of pedestrian exposure in a housing area might be defined in terms of some function of the number of pedestrians and their age and sex distributions together with the amount and character of the traffic and certain geometric characteristics of the layout of the roads and adjacent buildings; the supposition is that these are the factors that appear likely to influence risk of accidents to pedestrians.

But at this time little is known about these relationships or their actual influence on pedestrian casualty rates. Apart from the inherent complexity of the subject, problems arise in the very collection of suitable data on levels of pedestrian exposure to risk. Because of the great lengths of roads in housing areas and because of the generally small numbers of both pedestrians and annual casualties per unit area, the collection of enough data requires large resources of manpower in the field. In 1970 some experiments were carried out at Imperial College to test the feasibility of carrying out surveys of housing areas on a large scale by using photographic or video apparatus from a moving car.

OBJECTIVES

On completion of these preliminary trials, sample housing areas throughout England and Wales were surveyed with the aim of determining levels of pedestrian activity and risk and to provide enough data on the various relevant factors to determine the importance of each in relation to pedestrian casualty rates (1).

The task was seen first as one of collecting and tabulating information on pedestrian activity, traffic flows, and land use and layout characteristics; population and other socioeconomic characteristics; and pedestrian casualty statistics. These data should enable typical and atypical conditions encountered in housing areas to be described and should make up a comprehensive data base for use in further studies. But beyond this, an important aim of the study was to try to develop models that explain the relationships between pedestrian exposure levels and casualty rates in different housing areas.

WORKING ASSUMPTIONS

The whole study was based on the working assumption that casualty rates can be explained mainly in terms of three sets of data: (a) those relating to population and socioeconomic characteristics as derived from census material; (b) those on land use and layout characteristics as derived from maps and observation; and (c) those on activity characteristics as derived from relatively short-period surveys of pedestrians and traffic in sample housing areas.

The data available for use in the study appeared likely to have certain shortcomings in that the time periods for the casualty, census, and pedestrian sets of data were not strictly compatible with each other. It was also realized that for any particular survey area, the absolute number of casualties per year was likely to be low. A further working assumption was adopted that given both a consistent basis for the collection of each separate set of data and examination of a large number of sample housing areas, it should be possible to produce useful models that help to explain the factors that influence the known variations in casualty rates.

SCOPE OF REPORT

The general procedure adopted in the surveys is outlined in the next section. Next the data for each of the regions of England and Wales are used to determine which of the main factors were significant in relation to the casualty rates. Then all the data are treated as a single data set to allow various stratification tests to be carried out and multivariate casualty models are described that use the census data only, the pedestrian and traffic data (the activity variables) only, and the layout data only. A final model is developed by using all these types of explanatory variables. The procedure adopted to test the validity of the models is described: The entire data set was subdivided into two equal parts, models were developed by using only one of the two subsets, and then these models were used to compare predicted casualty rates with the actual rates for the second of the two subsets. In the last section the factors influencing the size of the predictive errors are examined and possible practical applications of the models are discussed.

METHODS AND DATA BASES

General Approach

The basic unit for the selection of sample survey areas was 1-km squares of the national grid. For any given developed square, the resident population can be obtained from the 1971 census, and the yearly number of pedestrian casualties can be obtained from police reports on road traffic accidents.

The 9 standard regions of England and Wales were adopted for the study, but the South East region was subdivided into 3 parts, making 11 regions in all (Figure 1). For each of these regions a sample of 40 to 50 squares of 1 km was selected, making a total sample of 474 squares (Figure 2). Data on land use, layout, population, socioeconomic characteristics, and casualties to pedestrians were collected; the sample areas were then surveyed to collect data on traffic and pedestrian numbers and age and sex distributions.

Selection of Sample Survey Areas

For each region, the universe from which sample survey squares were chosen was the set of 1-km squares containing parts of the built-up areas of all towns whose 1971 populations exceeded 1,500. But all squares with less than 20 percent of the area developed and all predominantly nonresidential squares (i.e., central areas and major industrial areas) were omitted.

The two criteria used for the selection of survey squares were the 1971 census population and the mean (1970-1971) annual number of pedestrian casualties in each square. The mean annual casualty rate per 100,000 population for each square was calculated and an ordered listing of these rates prepared. Every fourth square was then selected to obtain the sample. This sampling approach had the practical merit of being relatively simple and applicable to all the 11 regions. It resulted in the selection of a set of sample squares having representative casualty rates per 100,000 population.

Layout and Land Use Data

A 1:2,500 scale map of each sample square was used in the field for navigational purposes and for recording various field observations relating to land use. The maps were also used to measure the extent of the developed area within the 1-km grid square. The color coding on 1:50,000 scale OS maps was used to classify the roads into class A and B roads and

FIGURE 1 Regional boundaries for selection of survey squares and analysis of data.

FIGURE 2 Distribution of the 474 grid squares.

minor roads (coded C and D), whereas cul-de-sacs were classified separately as group E.

On the assumption that certain layout factors may play a role in determining accident rates or relative risk ratios or both, an attempt was made to develop quantified measures that could be used consistently to describe the salient physical characteristics of all types of development. However, it was accepted that it would not be possible to describe the layout characteristics in great detail (e.g., to record slight or irregular changes in street width from one end of a street block to the other). Accordingly, the following basic measures of layout were adopted:

1. Road lengths of types A, B, C, D, and others (E);
2. Numbers of junctions of various types;
3. Carriageway widths for each road type; and
4. Numbers of traffic signals, pedestrian crossing facilities, and bus stops.

Although most of the sample squares consisted of land in residential use, many had at least a small number of nonresidential uses such as shops or nursery schools, but some included shopping centers, large schools, or industrial areas. When such nonresidential uses are extensive, they may generate larger numbers (and different types) of pedestrians. It therefore is desirable to quantify these other uses so that the pedestrian exposure levels and casualty rates associated with the residential population and age and sex structure of each square can be determined. Data were therefore collected in the field on numbers or frontages of shops, and the presence of schools and other nonresidential uses were noted. A simple point systems was employed to define the amount of nonresidential land use in each square.

Census and Casualty Data

The data for each sample grid square were obtained from the 100 percent population and the 100 percent household census returns. These data were particularly useful in that they enabled resident populations and observed pedestrians to be compared in terms of their age and sex distributions.

Casualty data included details of age and sex of pedestrian casualties and other details such as class of road and pedestrian's activity at the time of the accident for the years 1969-1975. Because of the generally small and variable numbers of casualties per square per year, the 7 years of data for each square were converted into a mean annual rate for use in the subsequent analyses.

The Apparatus

For the field surveys, it was necessary to traverse all roads in each sample square, collect data on traffic and pedestrian numbers, and record details of land use and layout that could not readily be obtained by other means. A visual and audio recording was made with a portable 0.5-in. National Panasonic video tape recorder and video camera with zoom lens coupled to a convex mirror projecting through the roof of the survey car. Each videotape ran for only 35 to 40 min but that proved adequate for recording the survey runs on the roads of a 1-km square, even though the actual surveys generally took 1.5 to 2 hr to complete. The sound track was used to make verbal identification of road sections, locations of bus stops, and certain land uses such as shops, schools, and industry. Three such recording units were available for the surveys, and these were used in Renault 4Ls with opening roofs.

Traffic Surveys

The field surveys were conducted by making videotape recordings as the survey vehicle passed along all the roads in each 1-km square. If generally only one survey per street was made, the total time spent in any one section of the road network was comparatively short. But the reliability of traffic counts derived from the recordings would then be low, particularly in the common case of roads with low traffic volumes. Resources of time and manpower did not permit a large number of survey runs to be made down each street. It was therefore decided that the best approach was to make four runs on each of the A and B (major) roads but only one run on the C and D (minor) roads. Traffic on the E roads (cul-de-sacs) was not recorded, although number of parked vehicles was.

This approach reflected the view that for the C and D roads with generally lower amounts of traffic, pedestrian accidents are essentially random occurrences that are unlikely to be explained in terms of observed traffic flows. But class A and B roads generally have traffic volumes or densities that are likely to be more highly correlated with the data on casualties. Adoption of the moving-observer technique, when based on videotape recordings of the repeated runs on the A and B roads, enabled the number of moving vehicles to be counted and likely errors of the estimate to be calculated. Initially, the traffic data were expressed in terms of vehicle densities per kilometer of road.

Pedestrian Surveys

Number, age, and sex of pedestrians on both sides of each length of road were recorded in every 1-km square. Pedestrians were classified into six age groups: 0 to 4, 5 to 9, 10 to 15, 16 to 24, 25 to 59, and 60 and over. These divisions follow those used in the casualty data but differ slightly from the age grouping used in the 1971 population census. The surveys were carried out between 9:30 a.m. and 12:30 p.m. and 2:00 and 3:30 p.m. on weekdays and in the main during the school term. All surveys were carried out in dry weather. Pedestrian crossing facilities (such as zebras) were few, and the observable flows of pedestrians crossing at random locations were generally low. It did not prove possible to collect data on number of pedestrians crossing the roads.

PRELIMINARY ANALYSES

Data Base

The total (1971) population resident within the 474 squares was 1,669,000. Road length as measured off 1:2,500 scale maps was 3670 km. Pedestrian casualties (1969-1975) average 2,223 per year. The total urban area from which the sample squares were selected was made up of 8,006 squares of 1 km, with a total population of 27,973,000. Thus the sample set embraced 5.92 percent of the area and 5.97 percent of the population of the sampling universe.

Regions Compared

In Tables 1 and 2 the data from the regional samples and their relation to the regional sampling universes are compared in terms of their numbers of 1-km squares, populations, total casualties, population densities, and casualty rates. Individual regions had total sample areas ranging from 3.5 to 8.0 percent of the regional sampling universes. But for East Anglia, a 25 percent sample was taken.

Population samples ranged between 3.1 percent and 10.5 percent (East Anglia, 26.2 percent). Mean population density per square kilometer was 3,494 for the whole universe of grid squares and 3,521 for the 474 sample squares. In the individual regions densities ranged between 2,505 (South West) and 5,815 (Greater London) persons/km^2. Annual pedestrian casualty rates per 100,000 population were 154.5 for the universe of grid squares, and averaged 140.8 for the sample squares; regional figures ranged between 87 (East Anglia) and 213 (Greater London).

Table 3 shows that for a range of key variables (including population, pedestrians, and vehicle counts) the mean values per square differ between regions by a factor of up to 2. But for pedestrian

TABLE 1 Number of 1-Km Squares and Total Resident Population in Each of the 11 Regional Sample Sets

	Number of 1-Km Squares			Population (000s)			Population per Square Kilometer	
Region	Whole Region	Sample	Percent	Whole Region	Sample	Percent	Whole Region	Sample
East Anglia	228	57	2.5	661	173	26.2	2,897	3,037
North	531	42	9.9	1,568	138	8.8	2,953	3,295
Yorkshire and Humberside	1,031	43	4.1	3,326	154	4.6	3,226	3,553
North West	540	41	7.6	2,148	175	8.1	3,977	4,271
East Midlands	527	42	8.0	1,675	164	9.8	3,179	3,902
West Midlands	949	41	4.3	3,518	180	5.1	3,707	4,380
Wales	524	44	8.4	1,396	147	10.5	2,665	3,344
South West	512	36	7.0	1,283	88	6.9	2,505	2,436
Greater London	1,205	42	3.5	7,007	218	3.1	5,815	5,182
Outer metropolitan area	1,176	42	3.6	3,172	111	3.5	2,697	2,645
Outer South East	783	44	5.6	2,219	121	5.45	2,834	2,754
Total area	8,006	474	5.9	27,973	1,669	6.0	3,494	3,521

TABLE 2 Mean Casualty Rates for Each of the 11 Regional Sample Sets

	Total Pedestrian Casualties, 1970 and 1971			Pedestrian Casualties[a] per Year per 100,000 Population		Pedestrian Casualties[a] per Year per Square Kilometer	
Region	Whole Region	Sample	Percent	Whole Region	Sample	Whole Region	Sample
East Anglia	1,145	302	26.3	87	87	2.5	2.7
North	4,544	422	9.3	147	153	4.3	5.0
Yorkshire and Humberside	9,427	427	4.5	142	139	4.5	4.9
North West	7,154	529	7.4	167	151	6.1	6.5
East Midlands	4,585	474	10.3	137	145	4.3	5.3
West Midlands	10,048	513	5.1	143	143	5.3	6.3
Wales	4,093	421	10.3	147	143	3.6	4.8
South West	2,079	186	8.9	81	106	2.0	2.6
Greater London	31,540	882	2.8	225	213	13.0	10.5
Outer metropolitan area	6,811	254	3.7	107	114	2.9	3.0
Outer South East	5,016	288	5.7	113	119	3.2	3.4
Total area	86,442	4,698	5.4	155	141	5.4	5.0

[a]Casualties consist of 1971 and 1972 data only.

TABLE 3 Comparison of Regions: Mean Values per Sample Square of Selected Key Variables

Region	Population	Pedestrians	Annual Pedestrian Casualties	Moving Vehicles	Parked Vehicles	Road Length (km)	Percentage of Area Developed
East Anglia	3,037	92	2.5	32	149	7.3	64
North	3,295	147	4.9	26	137	8.2	51
Yorkshire and Humberside	3,553	95	4.2	31	156	8.3	54
North West	4,271	149	6.1	31	217	9.3	66
East Midlands	3,902	119	6.1	26	195	7.4	56
West Midlands	4,380	102	5.6	40	187	8.1	69
Wales	3,344	96	4.8	30	238	7.4	49
South West	2,435	85	2.6	25	169	6.9	50
Greater London	5,182	138	9.9	55	519	9.0	70
Outer metropolitan area	2,561	77	2.6	23	132	6.6	54
Outer South East	2,754	78	2.8	26	177	6.9	53
Total area	3,511	107	4.7	31	206	7.7	58

casualties the discrepancy widens: Greater London sample squares have almost four times as many pedestrian casualties as squares in East Anglia or the South West. Road length and size of the developed area, on the other hand, are relatively constant.

Definitions of various pedestrian casualty rates used in Tables 4 and 5 are given as follows:

- Cas: annual pedestrian casualty total per sample square,
- Cas/pop: annual pedestrian casualty rate per 100,000 residents,
- Cas/ped: annual pedestrian casualty rate per 100 pedestrians observed,
- Cas/veh: annual pedestrian casualty rate per 100 moving vehicles observed,
- Cas/dev: annual pedestrian casualty rate per square kilometer of developed area,
- Cas/km: annual pedestrian casualty rate per kilometer of road length,
- $Cas/(PV)^{1/2}$ annual pedestrian casualty rate per square root of the product of observed pedestrian and vehicle numbers on a sample square.

In Table 4 the regions are compared in terms of various alternative pedestrian casualty rates that were considered. By ranking the regions in terms of these different casualty rates, Table 5 shows that whatever the particular rate adopted, the order in the ranking is more or less the same: Greater London has the highest, and East Anglia the lowest, casualty rates.

Table 6 shows, for the 474 squares, the distribution of the pedestrian casualties in terms of severity, age, sex, time of day and year, and pedestrian's location and action at the time of accident. For the 474 grid squares as a whole, the number of pedestrians observed was about 3.0 percent of the

TABLE 4 Mean Pedestrian Casualty Rates for Samples Squares of Different Regions

Region	Cas/Pop	Cas/Ped	Cas/Veh	$Cas/(PV)^{1/2}$	Cas/Dev	Cas/Km
East Anglia	83	2.7	7.9	5.0	3.5	0.35
North	153	3.4	19.0	8.5	8.8	0.61
Yorkshire and Humberside	119	4.4	14.8	8.3	7.1	0.51
North West	142	4.1	19.5	9.3	8.9	0.65
East Midlands	156	5.1	23.6	11.6	9.3	0.82
West Midlands	130	5.6	14.0	9.2	7.4	0.69
Wales	142	5.0	16.0	9.3	8.9	0.64
South West	106	3.4	10.3	5.8	4.5	0.37
Greater London	191	7.2	18.0	12.2	12.4	1.10
Outer metropolitan area	102	3.4	11.2	7.1	4.5	0.39
Outer South East	102	3.6	10.9	6.6	4.6	0.40
Total area	134	4.4	15.0	8.7	7.2	0.61

Note: Pedestrian casualty rates are defined in the text.

TABLE 5 Rankings of Regions in Terms of Alternative Casualty Rates

	Cas/Pop	Cas/Ped	Cas/Veh	Cas	Cas/Km	Cas/Dev	$Cas/(PV)^{1/2}$
High casualty rate	9	9	5	9	9	9	9
↑	5	6	4	4	5	5	5
	2	5	2	5	6	4	4
	4	7	9	6	4	7	7
	7	3	7	2	7	2	6
	6	4	3	7	2	6	2
	3	11	6	3	3	3	3
	8	2	10	11	11	11	10
	10	8	11	8	10	8	11
↓	11	10	8	10	8	10	8
Low casualty rate	1	1	1	1	1	1	1

Note: Region 1, East Anglia; 2, North; 3, Yorkshire and Humberside; 4, North West; 5, East Midlands; 6, West Midlands; 7, Wales; 8, South West; 9, Greater London; 10, outer metropolitan area; and 11, outer South East.

TABLE 6 Key Characteristics of the Casualty Distribution for the 474 Sample Squares

Variable	Percentage of All Casualties (All Squares) Children (N = 7,809)	Adults (N = 6,942)	Total (N = 14,751)
Severity			
Fatal	1.3	4.9	3.0
Serious	28.0	31.3	29.6
Slight	70.7	63.8	67.4
Age (years)			
0-4	20.0	–	10.6
5-9	49.8	–	26.4
10-15	30.2	–	16.0
16-24	–	20.5	9.6
25-59	–	41.9	19.7
60+	–	37.6	17.7
Sex			
Male	59.6	49.3	54.7
Female	40.4	50.7	45.3
Time of day (hr)			
0-7	0.8	6.4	3.4
8-9	10.9	9.7	10.3
10-11	7.1	11.5	9.2
12-13	15.4	11.6	13.6
14-15	16.1	11.6	14.0
16-17	30.3	16.5	23.8
18-19	14.3	10.0	23.8
20-23	5.1	22.7	13.4
Month			
January	6.9	10.1	8.4
February	6.9	7.9	7.4
March	8.6	8.0	8.3
April	9.0	7.5	8.3
May	9.9	7.2	8.6
June	9.3	6.9	8.2
July	9.1	6.9	8.1
August	8.3	6.9	7.7
September	8.5	7.7	8.1
October	9.6	8.2	8.9
November	7.5	11.3	9.3
December	6.4	11.4	8.7
Pedestrian action			
Crossing at a pedestrian crossing	5.7	13.3	9.3
Crossing within 50 yd of pedestrian crossing	3.0	6.2	4.5
Pedestrian location			
In the road, not crossing	4.6	8.2	6.3
Masked by stationary vehicle	31.2	11.5	21.9
Daylight conditions	88.7	62.0	76.1
Junction type			
T or Y	36.0	39.3	37.6
X or multiple	13.0	16.5	14.6
Automatic signal control	4.3	8.2	6.1
Parked vehicle contributing to accident	17.9	7.0	12.8

resident population, the figure varying between 2.3 percent and 4.6 percent in the individual regions. The ratio of the child (0 to 15 years) casualty rate per 100,000 population of the same age group to the rate for the rest of the population varied in the regions between 1.8 and 5.3.

In Table 7 the data from the 474 squares are stratified into six age groups and the distributions of their resident population, observed pedestrians, and casualties are shown. Comparison of columns 2 and 3 with columns 3 and 4 shows that the 0-4 and the 10-15 age groups were less in evidence as pedestrians than in the resident populations. The highest casualty rate in column 8 was 428 annual casualties per 100,000 residents for the 5-9 age group, and the lowest was 62 for the 25-59 age group. If the casualty figures are expressed as rates per 100 pedestrians (column 9), the 10-15 age group had the highest rate. This is explained at least partly by the fact that the surveys were carried out mainly during school hours and consequently pedestrians 10 to 15 years old were underrepresented.

Key Variables and Simple Linear Regressions

In the preliminary regional analysis, mean values of key variables and simple linear correlations between selected variables and annual casualty rates per 1-km square were calculated (Table 8). Resident population per grid square kilometer gave correlation coefficients ranging from 0.66 to 0.92 for the 11 regions and 0.81 for all 474 squares. A number of other variables were found to have fairly high values of R. These included number of pedestrians (P) (0.60 and 0.89 for regions 10 and 9, respectively), moving vehicles (V) (0.65 and 0.85 for regions 6 and 8, respectively), $(PV)^{1/2}$ (0.76 and 0.92 for regions 2 and 11, respectively), and number of parked vehicles (0.58 and 0.95 for regions 10 and 11, respectively).

For the 474 squares as a whole, equations of the regression lines for the foregoing variables in relation to annual casualty rates per grid square kilometer were derived (Table 9). The standard errors of the estimate for the seven variables listed ranged from 3.8 to 5.3 casualties per grid square per year, and these may be compared with the overall annual mean of 4.67 casualties per square.

When resident population is used as the explanatory variable, all 11 regions have regression line slopes fairly similar to the slope for all 474 (national) squares, except for East Anglia and the outer metropolitan region. When number of observed pedestrians was used as the variable, the regional slopes lay close to the national slope, except for the previous two regions and also the northern and South West regions. In those four cases, there is a smaller rate of increase in number of casualties per grid square as number of pedestrians increases.

But the use of these rates per grid square may not provide the best basis, either for purposes of analysis or for practical application, of the derived casualty models. The data and the simple correlations were therefore examined in terms of the rates per square kilometer of developed area. On

TABLE 7 Distribution of Mean Resident Populations, Observed Pedestrians, and Casualty Rates by Age Group

Age Group (years)	Resident Population No.	Percent	Observed Pedestrians No.	Percent	Pedestrian Casualties No.	Percent	Casualties per 10^5 Population	Casualties per 100 Pedestrians
0-4	276	7.9	4.9	4.7	0.50	10.7	181	10.2
5-9	292	8.3	9.6	9.3	1.25	26.8	428	13.0
10-15	316	9.0	4.7	4.5	0.75	16.0	238	16.0
16-24	472	13.4	15.9	15.4	0.45	9.6	95	2.8
25-29	1,479	42.1	46.3	44.9	0.91	19.4	62	2.0
60+	676	19.3	21.9	21.2	0.82	17.5	121	3.7
All	3,511		103.3		4.68		133	4.5

TABLE 8 Correlations of Key Variables with Annual Pedestrian Casualties

Region	Population	Pedestrians (P)	Moving Vehicles (V)	$(PV)^{1/2}$	PV	Road Length (km)	Parked Vehicles
1	0.66	0.72	0.79	0.85	0.73	0.62	0.82
2	0.82	0.62	0.76	0.76	0.66	0.59	0.73
3	0.76	0.80	0.71	0.86	0.89	0.74	0.89
4	0.78	0.70	0.80	0.84	0.91	0.67	0.74
5	0.92	0.88	0.83	0.91	0.91	0.70	0.94
6	0.80	0.77	0.65	0.80	0.77	0.52	0.84
7	0.82	0.80	0.77	0.85	0.86	0.72	0.84
8	0.85	0.83	0.85	0.90	0.84	0.66	0.92
9	0.83	0.89	0.72	0.91	0.87	0.60	0.83
10	0.38	0.60	0.68	0.80	0.73	0.43	0.58
11	0.78	0.83	0.83	0.92	0.94	0.68	0.95
All	0.81	0.74	0.71	0.82	0.79	0.61	0.80

TABLE 9 Annual Pedestrian Casualties Related to Key Explanatory Variables

	Rates per 1-Km Square			Rates per Square Kilometer Developed			Rates per Kilometer of Road		
Explanatory Variables	Correlation Coefficient	SEE	Equation of Regression Line	Correlation Coefficient	SEE	Equation of Regression Line	Correlation Coefficient	SEE	Equation of Regression Line
Population	0.81	3.9	-3.46 + 0.0023X	0.69	6.0	-5.45 + 0.0022X	0.57	0.5	-0.22 + 0.00171X
Road length (km)	0.61	5.3	-6.08 + 0.0014X	0.26	7.9	-1.11 + 0.0006X	0.44	0.5	-0.13 + 0.000084X
Observed pedestrians (P)	0.74	4.5	0.08 + 0.0432X	0.59	6.7	2.20 + 0.0284X	0.54	0.5	0.23 + 0.0222X
Observed moving vehicle (V)	0.71	4.6	-1.37 + 0.1934X	0.54	6.9	-0.14 + 0.0139X	0.53	0.5	0.062 + 0.1144X
Observed parked vehicles	0.8	4.0	0.032 + 0.0226X	0.7	5.9	0.371 + 0.208X	0.7	0.4	0.046 + 0.01974X
PV	0.79	4.1	2.12 + 0.0005X	0.64	6.4	4.19 + 0.000267X	0.51	0.5	0.4 + 0.00176X
$(PV)^{1/2}$	0.82	3.8	-1.80 + 0.12X	0.73	5.7	-2.01 + 0.1038X	0.69	0.4	-0.049 + 0.0855X

Note: SEE = standard error of regression equations.

this basis (for the 11 separate regions), parked vehicles per developed square kilometer and $(PV)^{1/2}$ generally had the highest simple linear correlation coefficients, whereas the values of r based on resident population, number of pedestrians, or number of moving vehicles were generally relatively lower. Simple regression equations are set out along with the associated standard errors of the estimate for all the main variables in Table 9. For the 474 squares as a whole, values of R when numbers of observed pedestrians (P), moving vehicles (V), parked vehicles, and $(PV)^{1/2}$ are used as explanatory variables ranged between 0.54 and 0.73.

Surprisingly, these correlation coefficients were lower than was the case when casualty rates were expressed using the 1-km grid square as the areal unit of measurement. Possible explanations for this were that the extent to which a grid square is actually developed might be determined by other factors that have a bearing on casualty rate, such as distance of the square from town center or size of town. The preliminary analyses led to the conclusion that the correlations between annual casualty rates and a number of key explanatory variables were sufficiently high to make a more elaborate analysis worthwhile.

FURTHER ANALYSES

Casualty Models Based on Stratification of Data

Next, casualty rates were modeled by using only casualty data that matched the survey data (in terms of period of the day). Data from all 474 squares were stratified in terms of 24-hr and off-peak daytime (OPDT) casualties. Simple regression analyses produced three models of annual casualty rates per grid square using only OPDT casualties. These models had lower values of R and higher standard errors of the estimate (SEEs) than were obtained by using the full 24-hr casualty data, and it was concluded that stratification of the data by time of day did not improve the understanding of the relationships.

It was thought likely that different age and sex groups might have different casualty rates and that these rates might be differently distributed by time of day. Twenty-four-hour and OPDT models were derived with pedestrian numbers weighted to reflect the relative vulnerability of the six age and sex groups. But the correlation coefficients of both these weighted models were lower than the values obtained for either the 24-hr or OPDT data by using unweighted pedestrian numbers as the explanatory variable.

Further casualty models were derived by using both observed pedestrian numbers and resident populations as explanatory variables, stratified again by age groups: children, adults, and the elderly. For children (0-15 years), casualty rates correlated well (R = 0.83) with numbers of resident child populations; similarly for the 60+ age group (R = 0.73). But for the adult group (16-59), the correlation with resident adult population was only 0.70.

When the number of pedestrians (stratified into the same three groups) was used as the explanatory variable, the simple regression model was best for the adult pedestrian numbers (R = 0.79) but the correlations for the children and the elderly were much weaker (R = 0.43 and 0.60, respectively). It was concluded that prediction of annual casualties to children and the elderly is likely to be best when census population data are used, whereas for the adult population, use of observed pedestrian numbers gives the best results.

The casualty data were next stratified by road type (i.e., major and minor roads). Simple regression models were derived in terms of numbers of pedestrians, parked vehicles, and road length for each of the two road types. For both types the simple correlation coefficients were much lower and the SEEs higher than was the case for similar models based on all road types.

The data were then examined in greater depth by stratifying the set of 474 squares according to the values taken by some key variables and then comparing the characteristics of the subsets of squares thus formed. Squares with low casualty totals also tended to have low population and activity levels and relatively small amounts of urban development. Squares with high casualty totals tended to score high on all other counts. Where squares were less than 1 km from a town center, casualty rates per resident were some 50 percent higher than elsewhere, whereas the rates per pedestrian were some 25 percent lower than elsewhere.

Casualty rate patterns vary considerably according to the particular denominator chosen. It was concluded that by using crude casualty totals, the arbitrary position of the national grid lines has an undue and confusing influence on the figures. On the other hand, if the casualty figures are expressed as rates per unit of developed area, the effect of that arbitrariness can be eliminated. This approach was therefore adopted in the last stages of the analysis.

Multivariate Models of Casualty Rates

Multiple-regression analyses were carried out with casualty rates per square kilometer of development as the dependent variable and with three different sets of independent variables: the activity set (including numbers of pedestrians, vehicles, and parked vehicles), the census set (including population, number of households, etc.), and the land use set (including road length, number of shops, etc.). The best of these equations are discussed in the following.

Activity Set

The best activity set models were as follows:

$$C = -3.09 + 0.013PRK + 0.017P + 0.063V \quad R = 0.80 \quad SEE = 5.4 \qquad (1)$$

$$C = -3.13 + 0.012PRK + 0.075\,(PV)^{1/2} \quad R = 0.81 \quad SEE = 5.3 \qquad (2)$$

where

C = annual pedestrian casualties per developed square kilometer of housing area,
PRK = number of parked vehicles per developed square kilometer,
P = number of pedestrians per developed square kilometer, and
V = number of moving vehicles per developed square kilometer.

Model 1 has a multiple correlation coefficient of 0.80 (SEE = 5.4 casualties per year per developed square kilometer). PRK, P, and V per developed square kilometer provided significant contributions to the equation. In model 2, PRK and $(PV)^{1/2}$ proved significant and the value of R rose to 0.81 (SEE = 5.3).

Census Set

The best census set models were the following:

$$C = 2.02 + 0.0029Pop + 0.034SH - 0.0091C_o - 0.0078CT \quad R = 0.85 \quad SEE = 4.8 \qquad (3)$$

$$C = 2.91 + 0.023H - 0.0087C_o - 0.015DW - 0.0061CT \quad R = 0.84 \quad SEE = 4.9 \qquad (4)$$

where

Pop = resident population per developed square kilometer,
H = number of households per developed square kilometer,
SH = number of households in shared dwellings per 1,000 households,
C_o = number of cars owned per developed square kilometer,
CT = number of council tenants per 1,000 households, and
DW = number of dwellings per developed square kilometer.

Significant explanatory variables included resident population and number of dwellings, households, cars owned, and council tenants per 1,000 households (all expressed as rates per developed square kilometer). These two models could be used in appropriate cases to predict the annual pedestrian casualty rate per developed square kilometer, provided that the census data could be matched in temporal and areal terms to the housing area.

Land Use Set

The best model using only the land use set was the following:

$$C = -15.5 + 0.0077FR + 7.3 \times 10^{-6}T + 1.7EC + 0.93SC + 0.05J + 10.6DEV + 0.58RL \quad R = 0.73 \quad SEE = 6.7 \qquad (5)$$

where

FR = shop frontage in meters per developed square kilometer,
T = resident population of nearest town,
EC = employment code (0-3, index of nonresidential land use),
SC = school code (0-3, index of number and type of schools),
J = total number of junctions per developed square kilometer,
DEV = proportion of developed area in grid square kilometer, and
RL = total road length.

Significant variables were shop frontage, town size, extent of industrial and office uses, number of schools, number of road junctions, road length, and proportion of the grid square that was developed. Although the multiple correlation coefficient of the model is lower than those for models 1-4, it is apparent that the land use variables influence the pedestrian casualty rate to a considerable extent.

Full Data Set

Finally, a model using all three types of data sets was derived:

$$C = -2.37 + 0.0029Pop + 0.091V - 0.0085C_o - 0.0072CT \quad R = 0.88 \quad SEE = 4.4 \qquad (6)$$

Significant variables were resident population, number of moving vehicles, car ownership, and number of council tenants per 1,000 households (all expressed as rates per square kilometer of developed housing area). It should be noted that the number of pedestrians does not appear as a variable in the model because its introduction along with two other variables (numbers of major road junctions and households in shared dwellings) only marginally increased the value of R from 0.88 to 0.89 and slightly decreased the value of the SEE.

Tests of Models Using Data Subsets

To test the robustness or validity of the models, the data set of 474 squares was divided into two subsets. An examination of the means and standard deviations of the key variables showed that each subset had values similar to those for the full data set. New models were constructed that were similar in form and used the same variables as the original models. Their multiple correlation coefficients and SEEs were slightly lower than those resulting from the use of the full data set, as would be expected.

These new models were then used to predict casualty rates for the other independent data subset of 237 grid squares, and the results were compared with the actual casualty rates. The means of the residuals were close to zero, and the standard deviations of the residuals were similar to the SEEs of the models.

The sample grid squares of the 11 survey regions had mean casualty rates that varied considerably from region to region as did mean population, number of pedestrians, and so on, per grid square. It may be asked how far the models really explain these differences in casualty rates. Comparison of the recorded mean annual casualty rates for each of the 11 regions with the mean rates as predicted by models 2 and 6 showed that for each model more than half the residuals were less than 1.0 casualty per year.

The two regions with the greatest differences in mean annual casualty rate were East Anglia (region 1) and Greater London (region 9). The differences between the actual casualty rate and that predicted by models 2 and 6 were +1.4 and +1.0 for the East Anglia sample and +0.6 and -1.8 for the Greater London sample.

CONCLUSIONS

The Models

Casualty rate models 1 and 2, based on the activity data set, and the census data set models 3 and 4 all explain between about 66 and 72 percent of the variation about the mean. Model 6, which uses all types of variables, explains nearly 80 percent of the variation. It was concluded that the census models 3 and 4 gave better results in the prediction of annual numbers of casualties among children or the elderly, whereas the activity models 1 and 2 were more effective for the 16-59 age groups. Tests of the model by using half the data as an independent data set as described previously appeared to confirm that models 1-6 are robust. But the SEEs are disappointingly high.

Factors Influencing SEEs

The SEEs of models 1-4 and 6 ranged from 5.4 to 4.4 pedestrian casualties per year per developed square kilometer. These figures may be compared with the mean recorded value of 7.2 (with a standard deviation of 8.3). To some unknown extent, random variations in the annual number of casualties may be responsible for the size of the SEEs. For example, the sample grid squares had about 30 pedestrian casualties per square over the 7-year period. A typical square that because of its inherent characteristics should have had 30 casualties over the same period might simply because of chance have actually had anywhere between 20 and 40 casualties (95 percent confidence interval).

But a number of other contributory factors need to be considered. First, the method of selecting sample 1-km grid squares was based on examination of the universe of urbanized squares and their ranking by casualty rates per 100,000 resident population. The casualty data used for this (1971 and 1972) closely matched the population used (1971 census), but the selection of every fourth of the ranked squares of course could not ensure that the chosen housing area samples were representative as regards size of urban settlement, age or sex or socioeconomic structure of the resident population, population density, or type of layout arrangement of the squares. Moreover, the numbers of pedestrians coming into or passing though the survey squares or both could not be determined. Nor was it possible to count the numbers of pedestrians crossing the roads surveyed. Such crossing activity may be an important factor influencing casualty rates.

Tests of the representativeness of the samples for the East Anglia and West Midlands regions indicated that no special bias had been introduced in the selection of the sample squares. But no further tests were carried out of the representativeness of the squares in terms of their layout or other characteristics by which a housing area may be said to be representative. Admittedly, the sampling procedure adopted had the advantage not only that it was simple and straightforward but also that it made it possible to bypass the difficult problem of precisely defining the term "representative housing areas."

As to the amount of data collected for each region, it is considered that the scale was about right; had the number of squares surveyed been doubled it is unlikely that the regional models would have shown any significant improvements. The number of samples for each region was, however, too small to allow the various stratification analyses to be undertaken on a regional basis. But the amount of data available (for instance, for the age and sex stratifications), when based on all 474 squares, was adequate and there is nothing to suggest a need for a more extended data set.

Another problem that may have influenced the size of the standard errors arose because of the unavoidably differing dates of the 1971 census data, the 7-year casualty data, and the field surveys. The underlying relationships between number of pedestrians observed (in 1976 and 1977) and resident (1971) population may have been obscured in those squares in which fairly large population changes or new development had taken place since 1971.

An examination was made of the effect of excluding casualty data from outside of the field survey times, and it was concluded that the models could not be improved by such a stratification. In general, it must be accepted that, viewed on a global scale, the data collected were rather remote from the detailed and local factors leading to accidents, especially because the actual factors contributing to casualties (e.g., pedestrian activity and time of day) were not analyzed in the main modeling process.

One shortcoming of this study has been the uncertainty about the extent to which the surveys of number of pedestrians and vehicles were representative of long-term average local conditions. In particular, seasonal factors and school holidays may have had some influence on the effectiveness of the models. Short-term variations in the number of pedestrians (especially because total numbers per survey square were generally low) certainly take place;

thus the number of pedestrians observed for any particular square is not necessarily representative. But to the extent that these short-term variations occur randomly, the effects of the short survey times on the complete data set of 474 squares and on the resultant casualty models may not have been considerable. In fact, the SEEs of the activity-based models (1 and 2) were only marginally greater than the errors of models 3, 4, and 6, none of which used number of pedestrians as an explanatory variable. But it seems possible that had three or four more full surveys been carried out at each location, the fuller activity data might have resulted in marginally improved equations.

Finally, the standard errors of the models might have been lower had more time been spent on deriving and testing the best possible composite variables, and the mathematical form of the models themselves could perhaps have been improved. But it is considered that further work on the current data set along such lines would not be likely to produce dramatic reductions of the SEEs.

Practical Application of Models

It has been concluded that the data collected and the variables used in the resulting models provide a good basis for explaining the variation in annual casualty rates in housing areas. It remains to be considered whether the models are both suitable and sufficiently accurate to allow than to be put to practical use.

On the question of their suitability, models 1 and 2 require data inputs on number of pedestrian and other activity variables. Collection of such data in existing housing areas is of course possible, although tedious, but where alternative large-scale housing developments not yet built are being examined, no activity observations are possible. To use predicted number of pedestrians, for example, by invoking the simple linear regression models would be hazardous. No attempt has been made in this study to develop multiple-regression models of numbers of pedestrians but further analysis of the data from the 474 squares would almost certainly enable a useful pedestrian-number model to be derived.

Models 3 and 4, based on census data, would eliminate the need for collecting activity data but would require (a) reasonably up-to-date census data and (b) census data properly matched with the geographical boundaries of the housing area being examined. This certainly can raise problems. Where models 3 and 4 are to be applied to proposed alternative developments, much of the necessary data on the population, number of households, and so on, can be predetermined by the design for the area, but for some of the variables, such as number of cars per square kilometer of development, presumably some predictive model would have to be invoked.

The relatively poor performance of the land use variable models suggests that as a basis for practical application, the activity or census variable models are definitely to be preferred. Model 6, which incoporates all three types of variables, has the best performance of the group of models, but its practical and easy use may be limited because both census and activity data are required as inputs.

Model 6 has an SEE of 4.4 casualties per year per square kilometer of development. This SEE may be compared with the overall mean of 7.2 casualties per year per square kilometer. If the equation is used to predict the number of casualties likely to occur within a square kilometer of housing area, there would be a 95 percent confidence interval of ±8.6 casualties per year. If the predictive errors on different squares are statistically independent, a housing area of 18 km^2 would enable a prediction accurate to ±2 casualties per year per developed kilometer of housing area. To reduce the predictive error to ±1 casualty per year, an area of 75 km^2 would be required.

The models could be usefully adopted as a basis for comparison, for example, where a local authority (or indeed a residents' group) may be concerned about the number of pedestrian casualties in a particular housing area. The equations would show whether the area's casualty record is better or worse than what might be expected on the basis of the national sample. In this way, the models would help in identifying problem areas. This kind of use of the models is possible whatever the size of the area to be considered, although of course the statistical significance of any differences between the two casualty rates increases with the size of the housing area being examined.

Pedestrian Exposure to Risk

A brief discussion of the concept of pedestrian exposure to risk and of the extent to which the data and models measure the levels of exposure and risk follows. In the main study it has been assumed that the so-called activity variables, number of pedestrians and moving vehicles, represent the main factors relating to levels of exposure. The assumption has plausibility, because in the absence of pedestrians or moving vehicles or both, there could be no pedestrian casualties. The composite variables PV or $(PV)^{1/2}$ can be fairly regarded as measures of exposure because they explain much of the variation about the mean casualty rate (Table 8). Equations 1 and 2 use $(PV)^{1/2}$ as an explanatory variable, but the number of parked vehicles per developed square kilometer also appears as a significant variable. As Table 6 shows, parked vehicles contributed to 12.8 percent of all pedestrian casualties and to 17.9 percent of casualties to children. But it may be that this variable also acts as a proxy for levels of activity not measured by the short-period surveys of pedestrians or traffic or both.

The census models 3 and 4 result in somewhat better equations without including number of pedestrians or vehicles as explanatory variables. From this it can be inferred either that the census variables function as proxies for the activity or exposure variables (and that owing to the short survey periods for the pedestrian and vehicle counts, the proxy variables are therefore to be preferred) or that over and above the activity or exposure variables used in models 1 and 2, other compounding factors relating to certain of the land use and census variables also influence the casualty rates. This view appears to be supported by the superiority of model 6, which includes both activity and census variables. It seems likely that in the absence of number of pedestrians from the model, resident popuation serves as a proxy variable.

ACKNOWLEDGMENT

The work described in this paper was carried out under contract to the Road User Characteristics Division of the Safety Department of the U.K. Transport and Road Research Laboratory. The Accident Investigation Division of the Laboratory provided the necessary pedestrian casualty data in computerized form, and for that help the author is indebted. Thanks are also due to the Map Library of the Department of the Environment, from which the numerous maps needed for the study were obtained. The field surveys were carried out by a number of drivers and navigators, who also assisted in the tedious work of extracting the data and preparing it for analysis. In particular, the author would like to thank L. Frost, D. Warriner, J. Davies, M. Wagrel, and A. Dunford for their contributions. The regional and national analyses were carried out by J. Goldschmidt, M. Iqbal, and P. Wood.

REFERENCE

1. D.H. Crompton. Pedestrian Exposure to Risk in Housing Areas. TRRL SR 749. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1982.

Publication of this paper sponsored by Committee on Pedestrians.

Pedestrian Characteristics and Exposure Measures

RICHARD L. KNOBLAUCH, HENRY N. TOBEY, and EVELYN M. SHUNAMAN

ABSTRACT

The objectives of this research were to identify specific pedestrian trip-making characteristics, develop pedestrian exposure measures, and examine these trip-making characteristics and exposure measures relative to accident information in order to determine the relative hazardousness of various pedestrian characteristics and behaviors. A large-scale field study was conducted in five standard metropolitan statistical areas (SMSAs). A total of 12,528 person-hr were devoted to observing vehicles and pedestrians at a stratified random sample of locations in five SMSAs. Volume and activity data were recorded for 612,395 vehicles and 60,906 pedestrians. In addition, 20,147 pedestrians were coded by demographic characteristics and behavior. A total of 1,357 sites were measured, photographed, and described. Data on pedestrian trip-making characteristics and behavior are presented: who walks, where they walk, how they walk (or run), and when they walk. Pedestrian exposure is described in terms of the number of pedestrian-vehicle (PV) interactions. Exposure data are presented in terms of various pedestrian and site characteristics. Relative hazardousness was determined by comparing the exposure data with pedestrian accident data. The relative hazard associated with various site characteristics, pedestrian and vehicle characteristics, and pedestrian and vehicle actions is described.

Nearly one of every five traffic fatalities is a pedestrian. Pedestrian accidents account for 5 percent of all traffic accidents. The nature and extent of the pedestrian accident problem has been examined in many accident studies (1-3). However, for accident data to be meaningful, they should be compared with the experience of the nonaccident population, or the population at risk. This information on the population at risk is called exposure data. With the exception of some British and Australian studies (4-7), little is known about the nature of pedestrian exposure. This project reports on what pedestrians are doing when they are walking from place to place on public rights-of-way.

The results of an FHWA project on pedestrian risk exposure measures are described. The project had three major goals:

1. To identify pedestrian trip-making characteristics and behavior,
2. To determine characteristics of pedestrian exposure, and
3. To determine relative hazardousness of pedestrian behaviors, activities, and various situational factors.

RESEARCH PROCEDURES

A goal of the project was to develop a defensible national estimate of pedestrian behavior. To do this, it was necessary to observe pedestrians at a sample of locations that would allow the observed behavior to be developed into a national estimate. A series of random and stratified-random procedures was used to select the data-collection areas and the data-collection sites within those areas.

Site Selection

City selection was based on NHTSA's National Accident Sampling System (NASS), which provided a statistically sound sample with a properly developed weighting system. The NASS system consists of 10 strata of approximately equal size. Each stratum