Estimating the Daily Volume of Crossing Pedestrians from Short-Counts

I. HOCHERMAN, A. S. HAKKERT, AND J. BAR-ZIV

The main issue of this paper is whether the distribution of pedestrian volumes during the day can be generalized in a way similar to vehicular flow distribution, so that the daily average can be deduced from short counts. Hourly whole day counts in 86 urban locations were used to determine the typical daily distributions of crossing pedestrians in residential and central business district streets in Israel. For each hour, the standard deviation and coefficient of variation were calculated, as measures of variation among locations, to determine the best time of day for performing short counts. The standard deviation of the hourly rate was generally 1 to 3.5 percent of the total daily volume. The coefficient of variation was generally 20 to 50 percent of the average hourly volume. Hours during nonpeak periods and afternoon peak were identified with relatively low dispersion. These hours should be preferred whenever short counts are performed for the purpose of risk estimates. The average daily estimates can be improved by using counts of 2 or more hours. In addition, 135 daily counts in 15-min intervals were used to examine the adequacy of such short counts for estimating hourly flows in residential locations. The four 15-min counts in each full hour were regarded as four independent samples from the same distribution. Coefficients of variation (CVs) among the 15-min counts were calculated for each hour and averaged over locations. The CVs varied across hours but in each case were greater than 30 percent. It was concluded that 15-min counts should not be used for the estimation of hourly pedestrian flows in residential areas.

Data collection is the backbone of experimental research. Even the most sophisticated research tools can lead researchers astray when the data being analyzed are insufficient, biased, or inaccurate. The current study deals with the proper gathering of one type of data used in the field of road safety, namely pedestrian counts (PedC). PedC are used both in research and for design purposes. PedC are used to determine warrants for pedestrian facilities such as pavements, crossings, pedestrian signals, and underpasses and overpasses. In safety research, PedC are used together with vehicle counts to determine exposure and assess risk. The findings are used to identify pedestrian blackspots, to determine the efficacy of pedestrian facilities, or to evaluate the effect of measures aimed at pedestrian safety in before and after studies.

Most vehicular traffic counts are automated, and thus relatively simple and inexpensive. Pedestrian counting, however, does not lend itself to simple mechanical counting for several reasons:

- Unlike vehicles, pedestrian movements are not limited to a linear track;
- Pedestrians can change direction, halt, or retrace at any point; and
- Pedestrians may move in groups, sometimes very close together.

When dealing with safety, the relevant pedestrian activity is usually road crossing. This activity may cause an encounter between the pedestrian and a vehicle, thus exposing the pedestrian to the risk of an accident. The observation area may be a marked crossing, its surroundings, a midblock section, or an intersection.

Because manual counts by observers are expensive, the estimation of pedestrian flows is often based on short counts of 5 to 20 min duration. For warrants, these short counts are performed during peak hour and then expanded to represent the whole hour count. In safety evaluation studies, short counts may be used to represent exposure during a period of several years.

Most pedestrian counting procedures date back to the 1950s. In England, Russell (*1*) divided urban streets into sections of 30 to 100 m. Each section was counted twice for a period of 5 min as the counting team moved along the street. Counts were performed during the hours of 1030 to 1245 and 1400 to 1630. Older (*2*) increased the count period to 6 min. Neither investigator justified the duration of the counts. Regarding the counting hours, Older explained that he avoided peak hour conditions, which may exist in some sections and not in others.

Russell used short counts to define a measure of risk to pedestrians. Risk was defined as the number of accidents in 2 years divided by the pedestrian volume in 10 min. This measure is based on the hidden assumption that pedestrian exposure during the 2-yr period can be adequately estimated from short counts.

Mackie and Older (*3*) and Jacobs and Wilson (*4*) found that the distribution of pedestrians by place of crossing along the street for the short count was similar to that of the whole day. However, the adequacy of short counts as a measure of exposure for a period of several years was not dealt with.

Studies in other European countries [see Jorgensen and Rabany (*5*)] used two 6-min counts, quoting the aforementioned English studies. In the United States, the *Highway Capacity Manual* recommends counts of 15 min during peak hour, for the purpose of warranting sidewalks and crossings (*6*).

The question of how well short counts represent exposure for a whole year or more is complex. Generally, the average annual daily pedestrian volume (AADPV) is assumed to be a proxy measure of exposure. The validity of this assumption presents an interesting problem that is outside the scope of this paper.

Transportation Research Institute, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel.

Accepting this assumption, the next question is how accurate is an estimate of AADPV that is based on a short count. To answer this question, it is necessary to first consider the calculation of such an estimate. First, a short count is usually taken to represent the flow of a whole hour. An estimate of the hourly flow is achieved by multiplying the short count by a factor. Next, the hourly estimate is expanded to represent a whole day's flow. To determine the expansion factor it is necessary to know the daily distribution of crossing pedestrians. Finally, this daily flow is taken to represent a typical day during the study period. If seasonal variations exist, correction should be made for them. In summary

$$aadpv = C_{ij} \times K \times D_i \times S_j \quad (1)$$

where

$aadpv$ = an estimate of AADPV,
C_{ij} = short-count value in hour i and season j,
K = hourly multiplier: 60/minutes of short count,
D_i = daily expansion factor for hour i, and
S_j = seasonal correction factor for season j.

Given that the three sources of variation in $aadpv$, C_{ij}, D_i and S_j are independent, the variance of $aadpv$, var($aadpv$), can be expressed as

$$\text{var}(aadpv) = K^2 \times f[\text{var}(C_{ij}), \text{var}(D_i), \text{var}(S_j)] \quad (2)$$

Var(C_{ij}) measures the day-to-day variation in pedestrian flow counted during the same period of day. It is a function of the count duration, as well as of site-specific idiosyncrasies. Var(D_i) represents the deviation of the expected daily distribution at a specific location, from the mean distribution used to calculate the daily expansion factor, for the specific hour. Thus, var(D_i) is a function of the homogeneity of the locations used to calculate the daily distribution. Var(S_j) is a measure of the variation of the seasonality factor within the season and among locations.

There is little seasonality in traffic flow in Israel (*7*) and this seems to be true for pedestrian flows as well. Flow distributions do change during school vacation periods and of course on weekend days (Friday and Saturday), but these changes are better dealt with by calculating special distributions for these periods when necessary than by using seasonality correction factors. Thus, two major sources of variation remain—random fluctuations within location [affecting var(C_i)], and variations in daily distribution among locations [affecting var(D_i)]. Because these two sources of variation are independent, the variance of $aadpv$ can be expressed as follows (*8*):

$$\text{var}(aadpv) = K^2 \times [c^2 \times \text{var}(D) + d^2 \times \text{var}(C) + \text{var}(D) \times \text{var}(C)] \quad (3)$$

where $c = E(C)$ and $d \times E(D)$. (The subscripts were dropped for the sake of clarity.)

Equation 3 may be obtained by expanding $Y = C \times D$ in a Taylor series about its expected value, $c \times d$. It is an exact evaluation of var($aadpv$), as second derivatives are zero, and var($aadpv$) can be conveniently expressed in terms of the coefficients of variation (CV). (CV is defined as the standard deviation divided by the expected value.) If V is defined as CV^2, so that $V(X) = \text{var}(X)/E^2(X)$, then Equation 3 takes the form

$$\text{var}(aadpv) = K^2 \times c^2 \times d^2 \times [V(C) + V(D) + V(C) \times V(D)] \quad (4)$$

In situations where seasonality exists, Equation 4 can be readily generalized as follows:

$$\text{var}(aadpv) = K^2 \times c^2 \times d^2 \times s^2 \times [V(C) + V(D) + V(S) + V(C) \times V(D) + V(C) \times V(S) + V(D) \times V(S) + V(C) \times V(D) \times V(S)] \quad (5)$$

If the variances of C, D and S are small relative to their expected value, var($aadpv$) can be approximated by

$$\text{var}(aadpv) = K^2 \times c^2 \times d^2 \times s^2 \times [V(C) + V(D) + V(S)] \quad (6)$$

Both sources of variation in Equations 3 and 4 (C and D) are addressed in the present paper. However, because of lack of repeated counts for the same periods and locations on different days, random fluctuations are discussed only in the context of estimating hourly flows from 15 min counts. The use of such short counts for hourly estimates in residential locations is examined.

The main issue dealt with in this paper is whether the distribution of pedestrian volumes during the day can be characterized in a way similar to vehicular flow distribution, so that the daily average can be deduced from short counts. Hourly whole-day counts are used to determine the typical daily distributions of pedestrians in residential and central business district (CBD) streets in Israel and to determine the best time of day for performing short counts.

METHODS

Existing data from studies carried out at the Road Safety Center during the period 1970 to 1972 were used in this study. The data consisted of full day counts by hours, performed on weekdays, at 86 locations on urban streets. Fourteen locations were in CBDs and 72 on collector and arterial streets in residential neighborhoods. Most of the counts were performed during the hours 0700 to 2200; a few were 24-hr counts. Some of the counts were conducted at pedestrian crossings, some near crossings, and others at midblock sections. According to Jacobs and Wilson (*4*), the daily distribution of pedestrians is not seriously affected by the place of crossing. Thus, data for all sites were combined and analyzed together.

The hourly pedestrian volumes at each site were converted to percentage of total daily volumes. Thus, for each site a daily distribution of crossing pedestrians was calculated. The use of percentage distributions, rather then absolute numbers, was needed because of the wide range of hourly volumes in our data. For example, pedestrian volumes in non-CBD locations during peak hours ranged from 17 to 959 and in CBD locations from 233 to 4,671. This wide range reflects not only differences among streets but also differences in the width of the sections that were counted.

The average distributions for CBD sites and for residential sites were calculated, along with the hourly standard deviations (SD) and coefficients of variation (CV: defined as the SD divided by the mean). These measure the stability of the hourly share of total daily volume among sites for each hour. Thus the SD and CV serve as measures of stability of the different daily expansion factors, which are based on hourly counts. Periods with relatively small coefficients produce better estimates of the daily expansion factor, yielding better estimates of the daily volumes.

For many of the sites, the hourly volumes consisted of four 15-min intervals. Some of these sites were counted again after a few months. Counts of less than a full day were also used. In all, there were 135 daily counts by 15-min intervals. These 15-min counts were used to assess the adequacy of short counts to represent hourly volumes. Most sites were in residential neighborhoods. Counts were carried out during the main hours of pedestrian activity, between 0700 and 2200. The average of pedestrian volumes was relatively low: 8 to 10 pedestrians per 15 min.

The four 15-min counts in each full hour were regarded as four independent samples from the same distribution. The assumption of independence seems reasonable (*9*). The assumption of identical distribution implies no trend within the hour. This may not always be correct, especially during morning peak hour. During other parts of the day volumes are fairly constant. When trends do exist, the assumption of common distribution results in an overestimate of the variation.

SDs were calculated for each hour to measure the dispersion among the four 15-min counts and, thus, the reliability of an hourly flow estimate based on a 15-min count. To compare among different sites, the SD was standardized by dividing by the average of the four counts (m) to produce the CV.

RESULTS

Daily Volume Distributions in Residential Neighborhoods

Presented in Figure 1 are the average daily (24-hr) volume distributions in four residential neighborhoods in the city of Haifa. The distributions are fairly similar and display three peak periods: a morning peak between 0700 and 0800, a noon peak between 1200 and 1300, and an afternoon peak between 1600 and 1900.

The average daily distributions in two cities, Haifa and Givatayim, are compared in Figure 2. As counts in Givatayim were performed during day and evening hours only, both distributions are displayed and compared for this period. Again, a high degree of similarity can be observed between the two distributions. Therefore, all 72 residential sites were combined

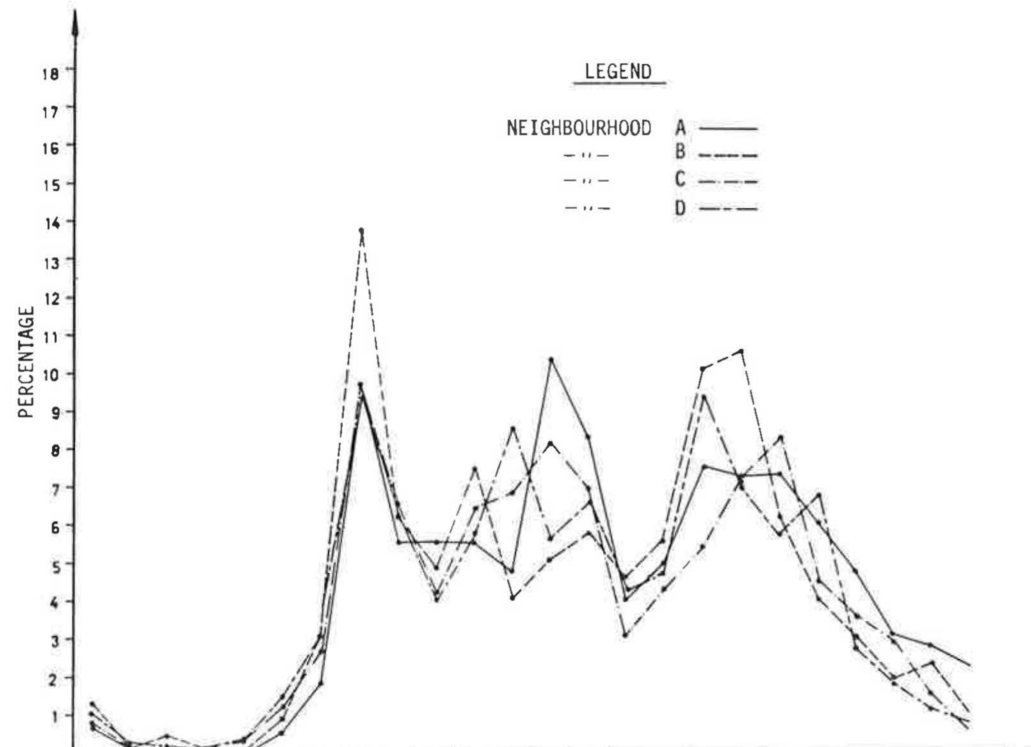

FIGURE 1 Daily distribution of crossing pedestrian volumes (percentages) in four residential neighborhoods in Haifa.

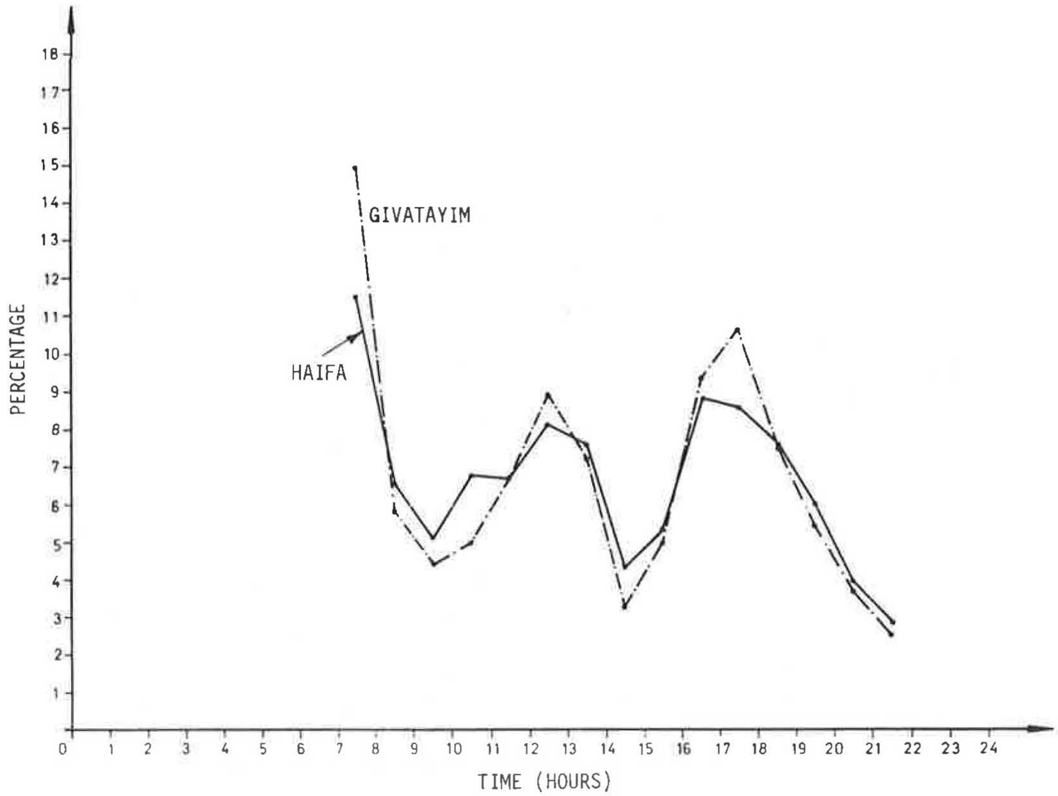

FIGURE 2 Comparison of daily distribution of crossing pedestrian volumes (percentages) in Haifa and Givatayim.

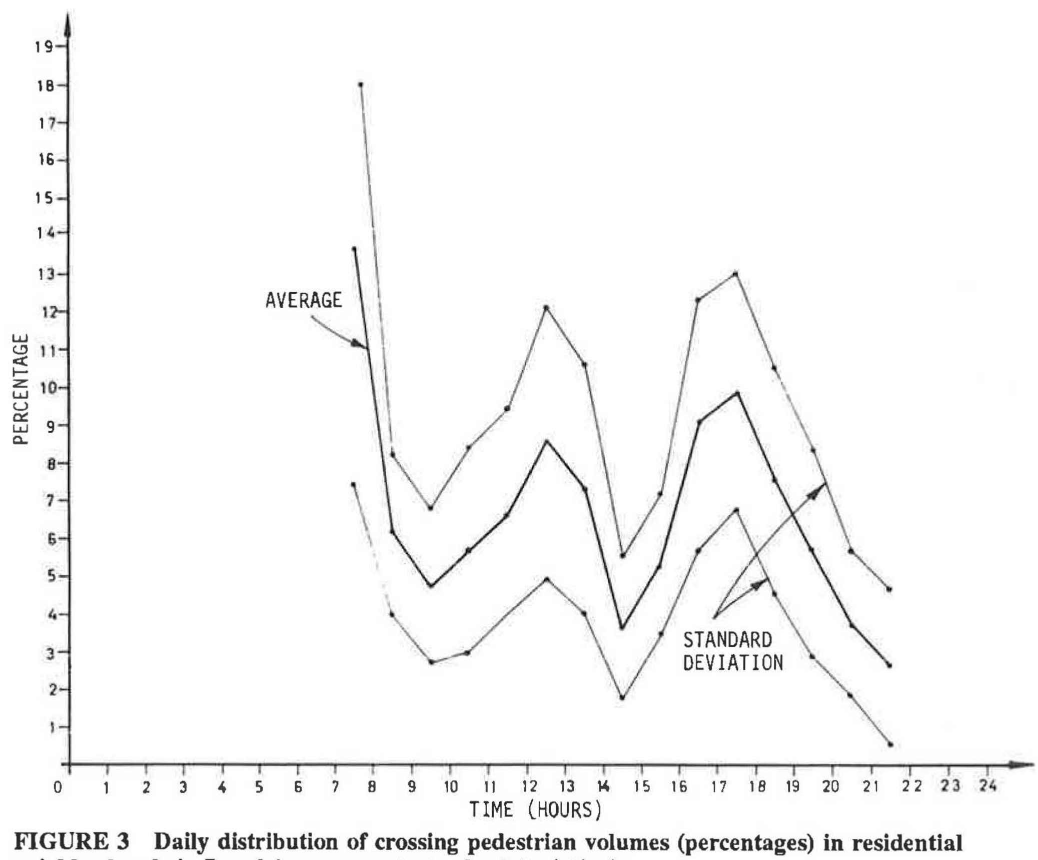

FIGURE 3 Daily distribution of crossing pedestrian volumes (percentages) in residential neighborhoods in Israel (average and standard deviation).

to calculate an average, or typical, distribution of pedestrian volumes in urban residential areas. This average distribution and the standard deviation of hourly pedestrian volumes are displayed in Figure 3. The same data are presented in Table 1, together with the hourly CVs.

The main results for residential areas are summarized as follows:

- Pedestrian volume during night hours, 2200 to 0700, constitutes only 3 percent of the total daily volume. Counts that do not include these hours are therefore almost complete. This result is based on the 24-hr counts performed in residential locations in Haifa.
- The daily distribution of crossing pedestrians has three peaks: a steep morning peak between 0700 and 0800, which carries 14 percent of the daily volume; an afternoon peak between 1600 and 1900, with an hourly flow of 8 to 10 percent of the total daily volume; and a midday peak between 1200 and 1300, containing 9 percent of the daily volume.
- The standard deviation of the hourly rate is 2 to 3.5 percent of the total daily volume, except for the morning peak, which has an SD of 6.1 percent. The lowest variation occurs during the nonpeak periods: 0800 to 1000 and 1400 to 1600, when the SD is about 2 percent.
- The CV is generally 30 percent to 50 percent of the average hourly volume (except for the hour 2100 to 2200, when pedestrian traffic is already very low). The lowest CVs occur again in the off-peak periods from 0800 to 0900 and from 1500 to 1600, when variance is low, but also during the afternoon peak from 1600 to 1900, when the variation in pedestrian volumes and the average flow are relatively high, so that their ratio remains relatively low.

TABLE 1 DAILY DISTRIBUTION OF HOURLY PEDESTRIAN CROSSING VOLUMES IN RESIDENTIAL AREAS

Hour	Mean (%)	SD (%)	CV (%)
7–8	13.6	6.1	0.45
8–9	6.2	2.1	0.34
9–10	4.7	2.1	0.44
10–11	5.7	2.6	0.46
11–12	6.6	2.7	0.40
12–13	8.6	3.7	0.42
13–14	7.4	3.4	0.46
14–15	3.7	1.9	0.52
15–16	5.3	1.9	0.35
16–17	9.1	3.2	0.35
17–18	9.9	3.2	0.32
18–19	7.6	3.1	0.40
19–20	5.6	2.8	0.49
20–21	3.7	2.0	0.53
21–22	2.6	2.2	0.82

NOTE: Sample size = 72 sites.

Daily Volume Distributions in CBD Areas

The average daily distribution of pedestrian crossing volumes in 14 CBD locations for the period of activity, from 0700 to 2200, is shown in Figure 4. The same information is presented together with the hourly SDs and CVs in Table 2. In addition,

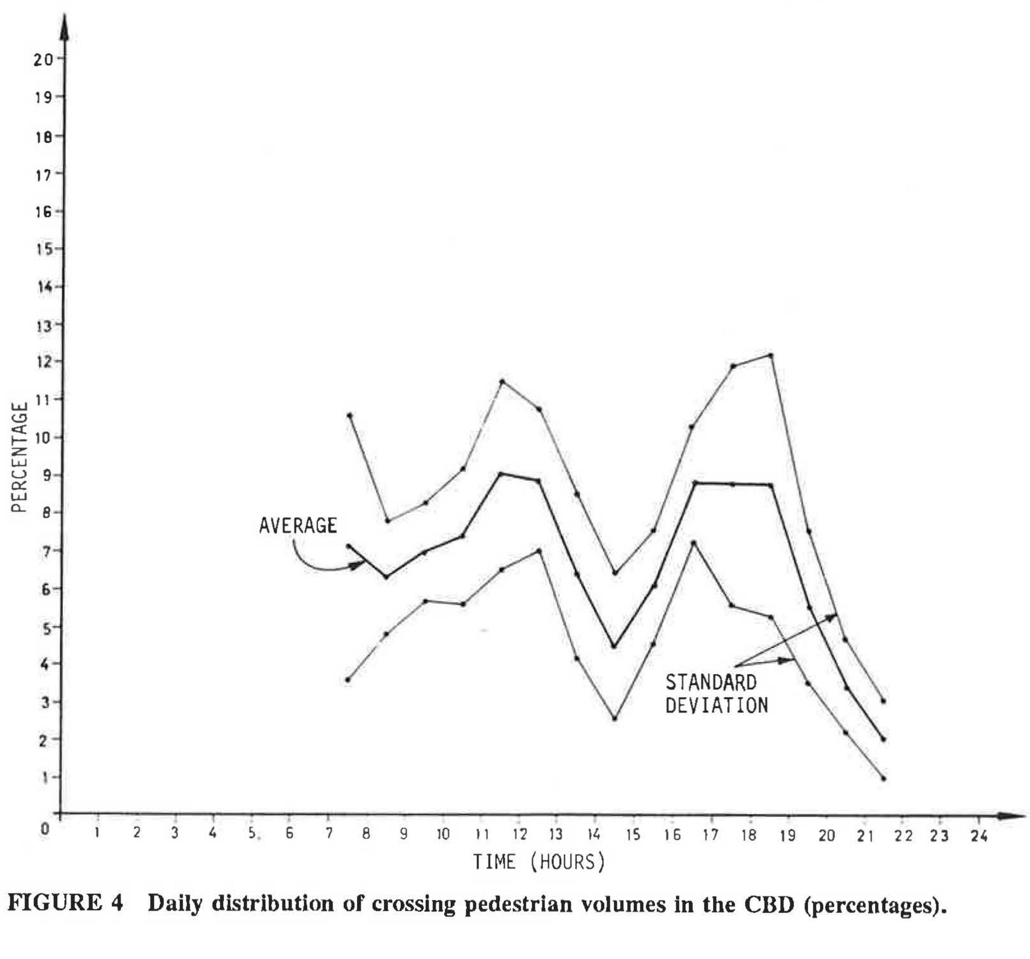

FIGURE 4 Daily distribution of crossing pedestrian volumes in the CBD (percentages).

TABLE 2 DAILY DISTRIBUTION OF HOURLY PEDESTRIAN CROSSING VOLUMES IN THE CBD

Hour	Mean (%)	SD (%)	CV (%)
7–8	7.1	3.5	0.49
8–9	6.3	1.5	0.23
9–10	6.9	1.4	0.19
10–11	7.4	1.8	0.24
11–12	9.1	2.5	0.28
12–13	8.8	1.9	0.22
13–14	6.5	2.2	0.34
14–15	4.5	1.9	0.42
15–16	6.1	1.6	0.25
16–17	8.8	1.5	0.17
17–18	8.8	3.1	0.35
18–19	8.8	3.5	0.39
19–20	5.6	2.1	0.37
20–21	3.6	1.3	0.37
21–22	2.1	1.0	0.47

NOTE: Sample size = 14 sites.

TABLE 3 ACCURACY OF SHORT 15-MIN COUNTS AS AN ESTIMATE OF THE HOURLY VOLUME

Hour of Counting	CV	No. of Sites	Ranking of Estimate Accuracy
7–8	51.8	40	9
8–9	66.9	57	13
9–10	67.7	62	14
10–11	37.5	31	2
11–12	48.0	63	8
12–13	55.2	70	12
13–14	53.1	64	10
14–15	45.7	18	6
15–16	32.7	28	1
16–17	41.2	67	4
17–18	45.3	64	5
18–19	46.0	57	7
19–20	40.1	41	3
20–21	53.8	45	11
21–22	69.1	46	15

data on pedestrian activity in the CBD during night hours were also available, but are not displayed here, to facilitate comparison with the daily distribution in residential areas.

The main findings for CBD areas are summarized as follows:

- Pedestrian volume during night hours, 2200 to 0700, constitutes only 7 percent of total the daily volume. Counts that do not include these hours are therefore almost complete.
- The daily pedestrian distribution in CBD locations is similar to that in residential locations and also contains three peak periods: morning (0700 to 0800), noon (1100 to 1300), and afternoon (1600 to 1900). The main difference between the two distributions is that the morning peak in the CBD is not as steep and contains only 7 percent of the daily volume. This difference reflects the fact that there are no schools in CBD areas (schools in Israel start between 0700 and 0800), whereas shops open only at 0900.
- The hourly variation is generally lower in CBD locations than in residential areas and amounts to 1 to 3.5 percent of the daily pedestrian volume. The variation is higher at peak hours than at off-peak periods, in accordance with the findings for residential sites.
- The CVs vary between 20 and 50 percent of the average hourly volume. CVs do not exceed 30 percent for the periods from 0800 to 1300 and 1500 to 1700, and they are lowest (<20 percent) for the hours from 0900 to 1000 and 1600 to 1700.

Short Counts

CVs for 15-min counts for each hour are presented in Table 3. The CVs are averaged over locations and range between 33 and 69 percent. The largest variations stem from trends within the hour, but even for hours with constant flows of pedestrians, the variation among short counts within the same hour is 33 to 50 percent. These findings are in accordance with those of Haynes (*9*), who recommends full hour counts when pedestrian flows are less than 10 pedestrians per minute.

SUMMARY AND DISCUSSION

Analysis of the daily distribution of pedestrian crossing volumes leads to a number of conclusions with practical implications for improved pedestrian counting. These can be used in the design of crossing facilities and for the study of pedestrian risk and safety.

- Typical daily distribution of pedestrian crossing volumes were developed for both residential and CBD locations in urban areas in Israel. These distributions can be used to estimate daily pedestrian volumes during the main hours of activity, 0700 to 2200. Total 24-hourly volumes can be estimated from 15-hourly volumes by multiplying by 1.03 for residential sites and by 1.07 in the CBD.
- The daily distribution for Israel was compared with similar distributions obtained for Germany and Australia. Neumann (*10*) conducted detailed counts at 111 sites in various German cities. Cameron (*11*) counted pedestrian volumes at 29 sites in Melbourne, Australia. The respective daily distributions are plotted in Figure 5. It can be seen that the three distributions have similar shapes, with differences reflecting differing lifestyles (e.g., school, work, and shopping hours).
- The standard deviation of the hourly rate in residential areas is generally 2 to 3.5 percent of the total daily volume. The hourly variation is generally lower in CBD locations and ranges between 1 and 3.5 percent of the daily pedestrian volumes. The lowest SDs occur during nonpeak periods. The coefficient of variation in residential areas is generally 30 to 50 percent of the average hourly volume. CVs smaller than 35 percent occur during the hours between 0800 and 0900 and 1500 and 1800. In CBD areas the CVs vary between 20 and 50 percent of the average hourly volume. CVs do not exceed 30 percent for the periods 0800 to 1300 and 1500 to 1700. These hours should be preferred whenever short counts are performed for the purpose of risk estimates. Counts during peak hours are still needed if facility design is the major aim. The average daily estimates can be improved by using counts of 2 or more hours. Tables 1 and 2 can be used to calculate the resulting SDs and CVs of such counts in Israel.

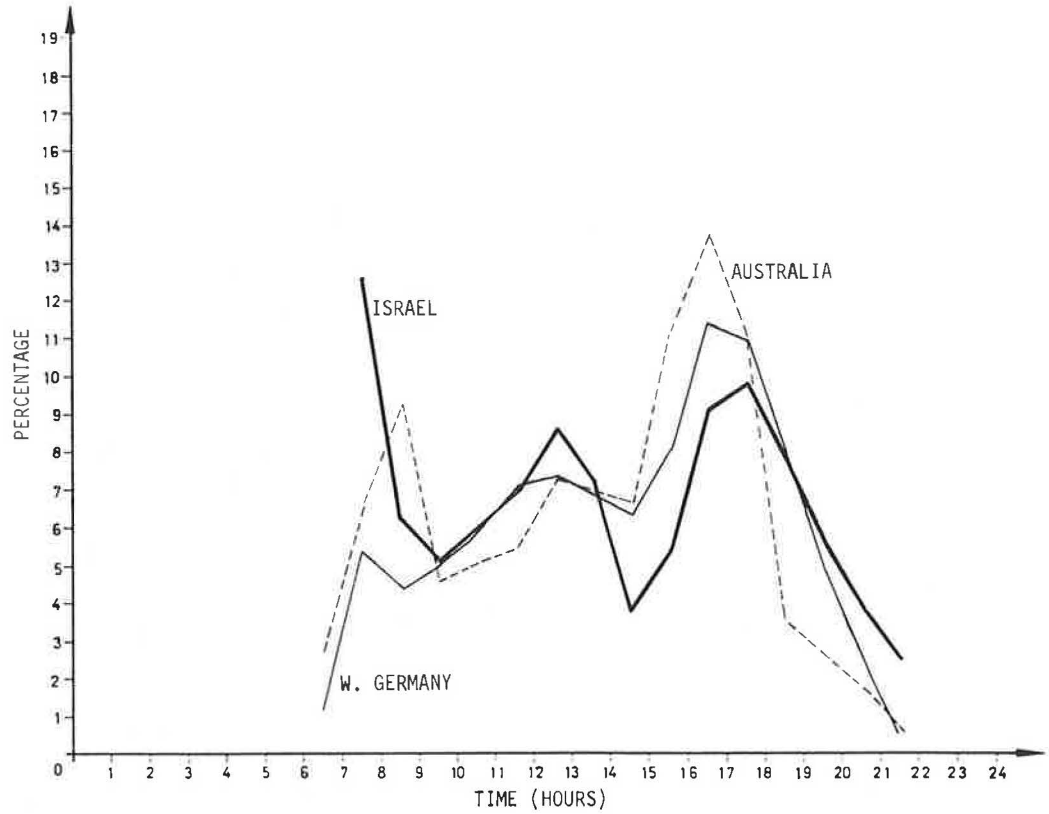

FIGURE 5 Comparison of daily distribution of crossing pedestrian volumes (percentages) in Israel, Germany, and Australia.

• Haynes (*9*) studied the accuracy of short pedestrian counts. Pedestrians were counted at 11 crossing sites in Norwich, England, with varied levels of activity. Counts were made during 1-min intervals for 2 hr. The data were then cleared of trends. Trend-cleared counts at 1-min intervals were found to be nearly normally distributed and independent of each other. The variance was found to be proportionate to the mean, which indicates that the accuracy of estimates based on these counts increases as pedestrian volumes increase. Assuming normality, Haynes developed a series of curves for various levels of activity. The relationship between maximum expected sampling error and length of count is shown for various levels of pedestrian activity in Figure 6. This figure can aid in choosing count duration to comply with a desired level of accuracy. It can be seen that for volumes of 30 pedestrians or more per minute, an error of less than 10 percent can be achieved with

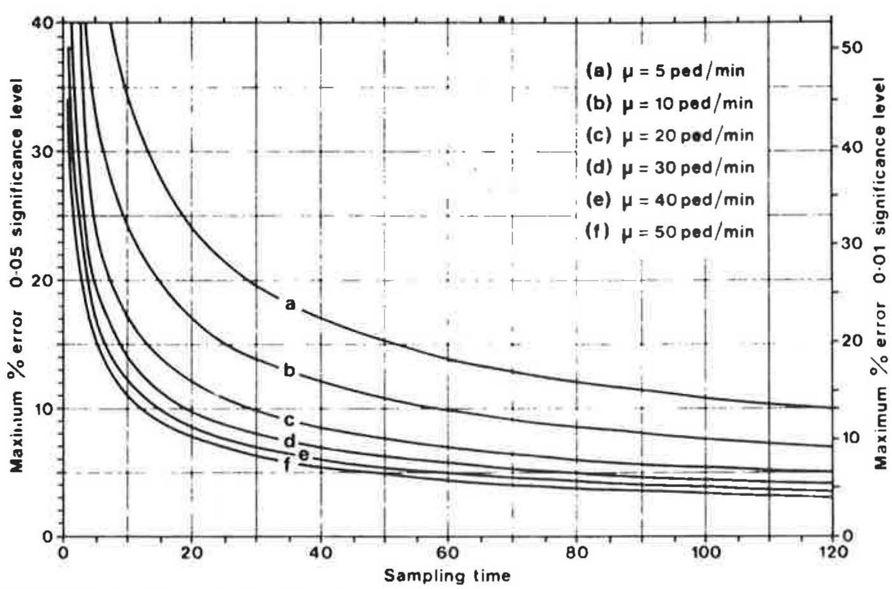

FIGURE 6 The relationships between maximum expected sampling error and sampling time for various levels of pedestrian activity (*9*).

12- to 15-min counts. For levels of 10 pedestrians per minute, full-hour counts are needed.

- Similar analysis on 15-min data at residential sites with relatively low pedestrian volumes was conducted in this study. The CV was found to change with the hours but in each case was greater than 30 percent. It is concluded that 15-min counts should not be used for the estimation of hourly flows in residential areas. When used to estimate daily flows, the short-count error is compounded to the error in the hourly expansion factor.

Pedestrian counts of varying durations are used to estimate AADPV, which serves as a measure of exposure. Further research on the day-to-day variation in hourly or daily counts is needed to fully address the issue of accuracy of these estimates. This will enable the estimation of $\mathrm{var}(aadpv)$ according to Equation 3.

Although the data used in the study date back to the early 1970s, it is not expected that pedestrian distributions have changed because basic life-styles in Israel have not changed during this period. However, this assumption should be checked periodically and the distributions updated if necessary.

REFERENCES

1. C. A. F. Russell. Relative Risk to Pedestrian or Crossings and Elsewhere. RN-1301-CAFR. Road Research Laboratory, Crowthorne, England, Jan. 1950.
2. S. J. Older. *A Survey of the Risk to Pedestrians Crossing Seven Roads in West London*. RN-4067-SJO. Road Research Laboratory, Crowthorne, England, Oct. 1961.
3. A. M. Mackie and S. J. Older. *A Study of Pedestrian Risk in Crossing Busy Roads in the Inner Suburbs of London*. LN 720. Road Research Laboratory, Crowthorne, England, Dec. 1964.
4. G. D. Jacobs and D. G. Wilson. *A Study of Pedestrian Risk in Crossing Busy Roads in Four Towns*. Report LR 106. Road Research Laboratory, Crowthorne, England, 1967.
5. W. O. Jorgensen and Z. Rabani. *Fodgaengerees sikkerhed i og ved fodgaengerovergange*. Report No. 7. Radet for Trafiksikkerhedsforskning, Copenhagen, Denmark, 1971.
6. *Special Report 209: Highway Capacity Manual*. TRB, National Research Council, Washington, D.C., 1985.
7. Y. Gur, A. S. Hakkert, and I. Hocherman. *Optimization of Traffic Counts in Israel* (in Hebrew). Report No. 85-84. Transportation Research Institute, Technion, Israel, 1985.
8. L. A. Goodman. On the Exact Variance of Products. *Journal of the American Statistical Association*, Vol. 55, No. 708, 1960.
9. R. Haynes. Sampling Time and Sampling Error in Pedestrian Counts. *Traffic Engineering Control*, Vol. 18, No. 2, London, England, Feb. 1977.
10. L. Neumann. *Further Studies on the Criteria for the Installation of Pedestrian Crossing Facilities*. BASt Project Report FP 8325. Bundesanstalt für Strassenwesen, Bergisch-Gladbach, Federal Republic of Germany, 1986.
11. M. H. Cameron. *Accident Risks at Zebra Crossings*. Australian Road Research Board (ARRB), Nunawading, Victoria, 1969.

Publication of this paper sponsored by Committee on Pedestrians.