

Hauer et al

We acknowledge and thank Dr. Hauer for allowing us to share this excellent and user-friendly tutorial with all the participants of the 2009 National SPF Summit, Chicago, Illinois.

Estimating Safety by the Empirical Bayes Method: A Tutorial.

Ezra Hauer, 35 Merton Street, Apt. 1706, Toronto, ON., Canada. Ezra.Hauer@utoronto.ca

Douglas W. Harwood, Midwest Research Institute, 425 Volker Blvd., Kansas City, MO 64110.

Dharwood@mrresearch.org

Forrest M. Council, Highway Safety Research Center, The University of North Carolina, Chapel Hill, N.C. , council@claire.hsrc.unc.edu

Michael S. Griffith, FHWA, Turner-Fairbank Highway Research Center, 6300 Georgetown Pike, McLean, VA 22101. Mike.Griffith@fhwa.dot.gov.

Abstract

The Empirical Bayes method addresses two problems of safety estimation; it increases the precision of estimates beyond what is possible when one is limited to the use of a two-three year history accidents, and it corrects for the regression-to-mean bias. The increase in precision is important when the usual estimate is too imprecise to be useful. The elimination of the regression to mean bias is important whenever the accident history of the entity is in some way connected with the reason why its safety is estimated. The theory of the EB method is well developed. It is now used in the Interactive Highway Safety Design Model (IHSDM) and will be used in the Comprehensive Highway Safety Improvement Model (CHSIM). The time has come for the EB method to be the standard and staple of professional practice. The purpose of this paper is to facilitate the transition from theory into practice

1. INTRODUCTION

The safety of an entity (a road section, an intersection, a driver, a bus fleet etc.) is “the number of accidents (crashes), or accident consequences, by kind and severity, expected to occur on the entity during a specified period.” (1, p.25). Since what is ‘expected’ cannot be known, safety can only be estimated, and estimation is in degrees of precision. The precision of an estimate is usually expressed by its standard deviation.

The safety of entities on which many accidents occur during a short period can be estimated quite precisely by using only accident counts. Thus, e.g., if on a road one expects 100 accidents per year, then, with three years of accident counts, one can estimate the average yearly accident frequency with a standard deviation of about $\sqrt{100/3} = \pm 5.7$ accidents/year or 5.7% of the mean. (This is based on the assumption that accident counts are Poisson distributed). Conversely, when it takes a long time for few accidents to occur, the estimate is imprecise. Thus,

Hauer et al

e.g., if one expects a rail-highway grade crossing or a driver to have one accident in ten years then, with three years of accident counts, the estimate of average yearly accident frequency has a standard deviation of $\sqrt{(0.1/3)} = \pm 0.18$. Since the mean is 0.1 accidents/year the standard deviation is 180% of the mean. Thus, one shortcoming of safety estimates that are based on accident counts only is that they may be too imprecise to be useful.

The other shortcoming of safety estimates that are based only on accident counts is that they are subject to a common bias. For practical reasons one is often interested in the safety of entities that either require attention because they seem to have too many accidents, or merit attention because they have fewer accidents than expected. In both cases, were one to estimate safety using accident counts only, the estimate would be biased. The existence of this 'regression-to-mean' bias has been long recognized; it is known to produce inflated estimates of countermeasure effectiveness. Yet, incorrect claims caused by failure to recognize this bias are still being published in the literature. (A recent example is, e.g., Datta et al. (2) who claim that low-cost treatments at three intersections in Detroit reduced total accidents by 44%, 48% and 57%. Yet, the three intersections were selected for treatment because their crash frequency, crash rate or casualty rate was higher than that of 95% of intersections and no correction for the regression-to-mean has been applied. Additional recent examples could be cited) Rational management of safety is not possible if published studies give rise to unrealistic expectations about the effectiveness of safety improvements.

The Empirical Bayes (EB) method for the estimation of safety increases the precision of estimation and corrects for the regression-to-mean bias. It is based on the recognition that accident counts are not the only clue to the safety of an entity. Another clue is in what is known about the safety of similar entities. Thus, e.g., consider Mr. Smith, a novice driver in Ontario who had no accidents during his first year of driving. Let it also be known that an average novice driver in Ontario has 0.08 accidents/year. It would be silly to claim that Smith is expected to have zero accidents/year (based on his record only). It would also be peculiar to estimate his safety to be 0.08 accidents/year (by disregarding his accident record). A sensible estimate must be a mixture of the two clues. Similarly, to estimate the safety of a specific segment of, say, a rural two-lane road, one should use not only the accident counts for this segment, but also the knowledge of the typical accident frequency of such roads in the same jurisdiction.

The theoretical framework for combining the information contained in accident counts with the information contained in knowing the safety of similar entities is the EB method. Starting with its application to road safety by Abbess et al. (3) the method is now well developed (1, Chapters 11 and 12) and has been widely applied. A recent application of the EB method of safety estimation is the Interactive Highway Safety Design Model (IHSDM, 4). Another application will be to the Comprehensive Highway Safety Improvement Model (CHSIM) now under development. The time has come for the EB method to be the standard of professional

practice; it should be used whenever the need to estimate road safety arises, whether in the search for sites with promise, the evaluation of the safety effects of interventions, or the assessment of potential safety savings due to site improvements. The purpose of this paper is to be the bridge between theory and practice.

2. THE EB PROCEDURE

The task is to make joint use of two clues to the safety of an entity: the accident record of that entity and the accident frequency expected at similar entities. This expected accident frequency at similar entities is determined by the Safety Performance Function (SPF) about which more will be said in section 3. In the EB estimate the joint use of the two clues is implemented by a weighed average. That is,

$$\text{Estimate of the Expected Accidents for an entity} = \text{Weight} \times \text{Accidents expected on similar entities} + (1 - \text{Weight}) \times \text{Count of accidents on this entity}$$

where $0 \leq \text{Weight} \leq 1$

The result is determined by how much ‘weight’ is given to the accidents expected on similar entities. The strength of the EB method is in the use of a ‘weight’ that is based on sound logic and on real data. This ‘weight’ will be seen to depend on the strength of the accident record (how many accidents are to be expected), and on the reliability of the SPF (how different may be the safety of a specific site from the average which the SPF represents).

The EB estimation procedure can be abridged or *full*. The abridged version makes use of the recent 2-3 years of accident counts and of the average traffic flow for that period. This reflects the now common belief that accident counts that are older than 2-3 years may not represent current conditions. However, the EB procedure removes most reasons for not using older data. Accordingly, the full version of the EB procedure makes use of a longer accident and traffic flow history. Because the full procedure uses more accident counts, the estimate of the full procedure is more precise than the estimate produced by the abridged procedure. Therefore, if data is available, one should strive to use the full procedure.

3. SAFETY PERFORMANCE FUNCTION AND WEIGHT.

The average accident frequency of ‘similar sites’ and the variation around this average are brought into the EB procedure by the Safety Performance Function (SPF). The SPF is an equation giving an estimate of μ , the average accidents/(km-year) for road segments or accidents/year for intersections, as a function of some trait values (e.g., ADT, Lane width, . . .) and of several regression parameters.

To illustrate, consider the SPF: estimate of $\mu = 0.0224 \times \text{ADT}^{0.564}$ for a certain kind of road in a given jurisdiction. Here ADT plays the role of one traits value, no additional trait values are

represented in the SPF, the estimate of one regression parameter is 0.0224, and the estimate of the second regression parameter 0.564. If on a road of this kind ADT=4000 vehicles per day, then one should expect $0.0224 \times 4000^{0.564} = 2.41$ accidents/(km-year).

SPFs are calibrated from data by statistical techniques. In the past it was common to assume that accident counts come from a Poisson distribution. However, researchers found that the accident counts used in the calibration of SPFs are usually more widely dispersed than what would be consistent with the Poisson assumption. This is why it is nowadays common to assume that the accident counts which serve as data come from a negative binomial distribution. One of the parameters of this distribution is the ‘overdispersion parameter’, denoted here by ‘v’. For road segments, the overdispersion parameter is estimated per-unit-length. That is, the dimension of v is [1/km] or [1/mile]. The meaning of v comes from the following relationship: if L is the length of a segment and 0 is the expected number of accidents for that segment, then the variance of accident counts on segments of that kind is $0[1+0/(vL)]$. The dimensions of v and L must be complementary. That is, if in the course of model calibration v is estimated per km, then L must be measured in kilometres. Note, v estimated per km = 0.622×v estimated per mile. For intersections L is taken to be one. More detail and an explanation of the sources of overdispersion is in reference (5)

Many SPFs and overdispersion parameters have been estimated and the results can be found in the literature. Thus, e.g., Maycock and Hall (6) model accidents at roundabouts, Hauer et al.(7) model accidents at urban signalized intersections, Bonneson and McCoy (8) model accidents at stop-controlled rural intersections, Miaou (9) models truck accidents on rural roads; Vogt and Bared (10) model accidents on rural road segments and intersections, Persaud and Dzbik (11) model accidents on freeways.

In summary we defined:

- μ the number of accidents/(km-year) for expected on similar segments and accidents/year expected for similar intersections.
- 0 the number of accidents during a specified period given by $\mu \times L \times Y$ expected for similar segments and $\mu \times Y$ expected for similar intersections. In this, L stands for segment length and Y for years.
- v overdispersion parameter estimated per unit length for segments. Naturally, entities for which the accident frequency is not proportional to their length (e.g. intersections or rail-highway grade crossings) have an overdispersion parameter that is not estimated per unit length.

It is now possible to give the expression for the ‘weight’ used in equation 1. In general:

$$\text{weight} = \frac{1}{1 + (\mu \times Y) / \phi} \quad \dots 2$$

where Y is the number of years of accident counts used. This expression for weight ensures that the variance of the estimate in equation 1 is as small as possible. For a full derivation and justification, see (1, pp. 193-194).

4. THE ABRIDGED EB PROCEDURE ILLUSTRATED.

To introduce the abridged procedure consider numerical examples of gradually increasing complexity:

Numerical Example 1: A Road segment with one year of accident counts.

A road segment is 1.8 km long, has an ADT of 4000, and recorded 12 accidents in the last year. The SPF for similar roads is $0.0224 \times \text{ADT}^{0.564}$ accidents/(km-year), with an overdispersion parameter $v=2.05/\text{km}$. To estimate the safety of this road segment proceed as follows.

Step 1: Average for entities of this kind.

Roads such as this have $0.0224 \times 4000^{0.564} = 2.41$ accidents/(km-year), on average.

Therefore segments that are 1.8 km long are expected to have $1.8 \times 2.41 = 4.34$ accidents in one year.

Step 2: Weight.

We need a 'weight' for joining the 12 accidents recorded on this road and the 4.34 accidents for an average road of this kind. For weight we use equation 2. Here $\mu=2.41$ accidents/(km-year), $Y=1$ and the estimate of $v=2.05/\text{km}$. Therefore: $\text{weight} = 1/[1 + (2.41 \times 1)/2.05] = 0.460$. Note that both μ and v are 'per unit length'.

Step 3: Estimate.

Using equation 1 the estimate of the expected accident frequency for the specific road segment at hand is: $0.460 \times 4.34 + 0.540 \times 12 = 8.48$ accidents in one year. Note that 8.48 is between the average for similar sites (4.34) and the accident count for this site (12). The EB estimator pulls the accident count towards the mean and thereby accounts for the regression to mean bias. The standard deviation of the estimate of the expected accident frequency is given by:

$$\sigma(\text{estimate}) = \sqrt{(1 - \text{weight}) \times \text{estimate}} \quad \dots 3$$

Here, $\Phi = \pm/(0.54 \times 8.48) = \pm 2.14$ accidents in one year.

Numerical Example 2: Three years of accident counts

Suppose now that for the same road segment we have three years of accident counts: 12, 7, 8, and that the ADT in each of those three years was 4000 vpd. To estimate the safety of the road segment:

Step 1: Average for entities of this kind.

As before, segments of this kind are expected to have 2.41 accidents/km-year. On 1.8 km in three years we expect $1.8 \times 3 \times 2.41 = 13.01$ accidents.

Step 2: Weight.

The weight is $1/[1+(2.41 \times 3)/2.05] = 0.220$. Note that with one year of accident data used the weight was 0.460. As more years of accident data as used, the weight (given to the number of accidents expected on similar entities) diminishes.

Step 3: Estimate.

Expected accidents $= 0.220 \times 13.01 + 0.780 \times (12+7+8) = 23.92$ accidents in three years with $\Phi = \pm / (0.78 \times 23.92) = \pm 4.32$ or $23.92 / (3 \times 1.8) \pm 4.32 / (3 \times 1.8) = 4.43 \pm 0.80$ accidents/(km-year).

Numerical Example 3: Application of Accident Modification Functions (AMFs)

Suppose now that the SPF equation in Example 1 is for roads with 1.5 m shoulders while the road segment of interest has 1.2 m shoulders, and that a 0.3m decrease in shoulder width is known to increase accidents by, say, 4%.

Step 1: Average for entities of this kind.

Using the result from Example 1, segments of this kind are expected to have $1.04 \times 2.41 = 2.51$ accidents/km-year. On 1.8 km in three years we expect $1.8 \times 3 \times 2.51 = 13.55$ accidents.

Step 2: Weight.

The weight is $1/[1+(2.51 \times 3)/2.05] = 0.214$.

Step 3: Estimate.

Expected accidents $= 0.214 \times 13.55 + 0.786 \times (12+7+8) = 24.12 \pm / (0.786 \times 24.12) = 4.35$ accidents in three years or $[24.12 \pm 4.35] / (3 \times 1.8) = 4.47 \pm 0.81$ accidents/(km-year).

Numerical Example 4: Subsections and Accident records.

Consider the road segment in Figure 1 that is made up of three subsections that differ in some traits (which determine the variable values of the SPF) and in the AMFs. However, the accident count is not available separately for each subsections, only for the entire 1.5 km segment on which 11 accidents were counted in the last two years.

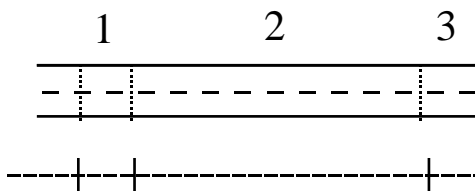


Figure 1

Step 1: Average for Entities of this kind.

The ADTs and AMFs differ amongst the subsections as shown in columns 2 and 4 of Table 1.

Table 1

Subsection	ADT	Length [km]	AMF	Accidents/(km-year)	Accidents
1	2000	0.1	.90	1.466	0.147
2	2300	1.2	.95	1.675	2.010
3	2300	0.2	1.05	1.851	0.370
Sum					2.527

Assume that, as in the earlier examples the SPF is $0.0224 \times \text{ADT}^{0.564}$ accidents/(km-year) and $v=2.05/\text{km}$. Thus, after correction for AMF, subsection 1 is expected to have $0.0224 \times 2000^{0.564} \times 0.90 = 1.466$ accidents/(km-year) and therefore $1.466 \times 0.1 = 0.147$ accidents/year. The three sub-sections together are expected to have $2.527 \times 2 = 5.054$ accidents in two years or $2.527/1.5 = 1.715$ accidents/(km-year). From here on it is convenient to forget about the subsections and treat the 1.5 km segment as one entity.

Step 2: Weight.

The weight is $1/[1+(1.715 \times 2)/2.05] = 0.374$.

Step 3: Estimate.

Expected accidents for the 1.5 km long section in two years $= 0.374 \times 5.054 + 0.626 \times 11 = 8.78 \pm (0.626 \times 8.78) = 2.34$ accidents or $[8.78 \pm 2.34]/(1.5 \times 2) = 2.93 \pm 0.78$ accidents/(km-year).

Numerical Example 5: Accidents by severity.

Consider again the setting in numerical example 2 with the addition of the information in columns 1 and 2 of Table 2.

Step 1: Average for entities of this kind.

As in the earlier examples, segments of this kind are expected to have 2.41 total accidents/(km-year). Applying the typical proportions in column 2 of Table 2, we expect 0.046 fatal accidents, 0.128 A-injury accidents, . . . , as shown in column 3. On 1.8 km in three years we expect on roads of this kind $1.8 \times 3 \times 0.046 = 0.247$ fatal accidents as shown in column 4.

Table 2

Accident severity	Accidents in three years	Proportion on similar roads	Average Accidents/(km-year)	Average Accidents in three years	Weight	Expected Accidents this site
	1	2	3	4	5	6
Fatal (K)	1	0.019	0.046	0.247	0.937	0.295
Incapacitating injury (A)	2	0.053	0.128	0.690	0.843	0.896
Non-incapacitating injury (B)	2	0.151	0.364	1.965	0.653	1.977
Possible injury (C)	5	0.140	0.337	1.822	0.669	2.872
Property damage only	17	0.637	1.535	8.290	0.308	14.317
Total	27	1.000	2.410	13.014		20.357

Step 2: Weight.

The weight for fatal accidents is $1/(1+0.046 \times 3/2.05) = 0.937$ as shown in column 5. The overdispersion parameter, v remains 2.05/km for all severities because it can be shown that when the SPF is multiplied by a constant, the overdispersion parameter is unchanged. Note that the weight of the 'Average for entities of this kind' is large for the rare accident severities. It is the property of the EB procedure that estimates will not be dominated by the random occurrence of rare events.

Step 3: Estimates.

The estimate of expected fatal accidents $= 0.937 \times 0.247 + 0.063 \times 1 = 0.295 \pm (0.063 \times 0.295) = 0.136$ accidents in three years. Note that the sum of expected accidents when estimated separately for each severity is 20.35. When the same has been estimated in example 2 using the total accidents without differentiation by severity, the estimate was 23.92 accidents. The discrepancy has two sources. First, it is appropriate that the specific accident severity of a site should be reflected in the estimates. Therefore, in principle, the two numbers should differ. However, there is a systematic reason for the discrepancy. It arises mainly because separation into severity classes inevitably results in smaller values of μ used in equation 2, and therefore in larger weights given to the expected accident frequency on similar entities. An ad-hoc correction could be to multiply each estimate by the ratio $23.92/20.35$. The estimate of expected fatal accidents would then be $0.295 \times 1.118 = 0.347$. A correct way of removing the blemish would be to adopt procedures described by Flowers (12) or Heydecker (13). However, both require additional parameter estimates to be used and these are, at this

August 2001

time, not easily available.

Numerical Example 6. An intersection.

For three-leg rural intersection in Minnesota Vogt and Bared (7) find that under nominal conditions μ is estimated by $6.54 \times 10^{-5} \times \text{ADT}_{\text{mainline}} \times \text{ADT}_{\text{minor road}}$ and the estimate of v is 1.96. Consider such an intersection with $\text{ADT}_{\text{mainline}}=4520$, $\text{ADT}_{\text{minor road}}=230$, the AMF to account for differences from nominal conditions is 1.27, and there were 7 accidents in three years.

Step 1: Average for entities of this kind.

Under the nominal conditions, intersections of this kind are expected to have $6.54 \times 10^{-5} \times 4520^{0.82} \times 230^{0.51} = 1.041$ accidents/year. Under the real conditions of this intersection, using the AMFs, $1.27 \times 1.041 = 1.322$ accidents/year. In the three years for which accident counts are used, $3 \times 1.322 = 3.966$ accidents.

Step 2: Weight.

The weight is $1/[1+(1.322 \times 3)/1.96] = 0.331$

Step 3: Estimate.

Expected accidents $= 0.331 \times 3.966 + 0.669 \times 7 = 6.00 \pm (0.669 \times 6.00) = 2.00$ accidents in three years or $[6.00 \pm 2.00]/3 = 2.00 \pm 0.67$ accidents/year.

Numerical Example 7. Accidents allocated to a group of intersections .

Some data bases contain information about how many intersection (and intersection-related) accidents have occurred on a road segment without the ability to specify how many occurred on which intersection. Consider a road segment with two intersections for which we have estimates of μ_1 (2.6 accidents/year), v_1 (2.2) and of μ_2 (4.3 accidents/year), v_2 (1.8). In three years, 11 accidents have occurred on these two intersections.

Step 1: Average for entities of this kind.

In the three years for which accident counts are available and on two similar intersections one should expect $3 \times 2.6 + 3 \times 4.3 = 7.8 + 12.9 = 20.7$ accidents.

Step 2: Weight.

Were one to use equation 2 directly, as if the two intersections were one, weight would be $1/(1+20.7/2) = 0.088$. In this the average overdispersion parameter was used. This is a bit of an oversimplification. Actually, when the accident count is available jointly for n entities with means $0_1, 0_2, \dots, 0_n$ and overdispersion parameters v_1, v_2, \dots, v_n and when correlation coefficient between 0_i and 0_j is $\Delta_{i,j}$ then the weight should be computed by:

$$\text{weight} = \frac{1}{1 + \frac{\sum_{i=1}^n \eta_i^2 / \varphi_i + 2 \sum_{i=1}^n \sum_{j=i+1}^n \rho_{i,j} \sqrt{\frac{1}{\varphi_i \varphi_j}} \eta_i \eta_j}{\sum_{i=1}^n \eta_i}} \quad \dots 4$$

But, it is at present not clear what correlation coefficient should be used and therefore the two extremes are of interest.

$$\text{When } \rho_{i,j}=0, \text{ weight} = \frac{1}{1 + \frac{\sum_{i=1}^n \eta_i^2 / \varphi}{\sum_{i=1}^n \eta_i}} \quad \dots 5$$

$$\text{When } \rho_{i,j}=1, \text{ weight} = \frac{1}{1 + \frac{\sum_{i=1}^n \sqrt{(\eta_i^2 / \varphi)^2}}{\sum_{i=1}^n \eta_i}} \quad \dots 6$$

In this example the weight is between $1/[1+(7.8^2/2.2+12.9^2/1.8)/20.7]=0.147$ and $1/\{1+[(7.8^2/2.2)+(12.9^2/1.8)]^2/20.7\}=0.085$.

Step 3: Estimate.

Using the simply-obtained weight of 0.088, Expected accidents= $0.088 \times 20.7 + 0.912 \times 11 = 11.94 \pm (0.912 \times 11.94) = 3.30$ accidents in three years.

5. THE FULL PROCEDURE ILLUSTRATED.

So far we discussed the abridged EB procedure. The full procedure differs from the abridged procedure in that year to year changes in ADT and in other variables can be brought into estimation thereby allowing use of longer accident histories. The full EB procedure is illustrated by numerical examples.

Numerical Example 8 - Accounting for changing ADTs

A road segment is 1.8 km long. It has remained physically unchanged during the past 9 years. The ADT estimates and accident counts for each year are given in rows 2 and 3 of Table 3. As in earlier examples, for this kind of road and nominal conditions μ is estimated by $0.0224 \times \text{ADT}^{0.564}$ accidents/(km-year) and the overdispersion parameter v is 2.05. Assume further that to convert from nominal to real conditions, the product of all AMFs is, in this case, 0.95. To estimate the safety of this road section in each of the nine years proceed as follows:

Table 3

1	Year	1989	90	91	92	93	94	95	96	97	Sums
2	ADT	4500	4700	5100	5200	5600	5400	5300	5300	5400	
3	Accidents	12	5	9	8	14	8	5	7	6	74
4	μ_{year} [accidents/(km-year)]	2.446	2.506	2.624	2.653	2.767	2.710	2.682	2.682	2.710	23.781
5	Expected accidents in year	4.402	4.511	4.724	4.776	4.980	4.879	4.828	4.828	4.879	42.806
6	Expected annual accident for segment	7.36	7.54	7.89	7.98	8.32	8.15	8.07	8.07	8.15	71.52

Step 1. Average for entities of this kind

Each year has an estimate of the expected number of accidents for roads of this kind. Thus, e.g., for 1989 and under nominal conditions, roads with ADT=4500 are estimated to have $0.0224 \times 4500^{0.564} = 2.574$ accidents/(km-year) and after adjustment to actual conditions $\mu_{1989} = 2.574 \times 0.95 = 2.446$ accidents/(km-year) as shown in row 4. Listed in row 5 are the expected accidents when segment length has been accounted for.

Step 2. Weight.

The formula for computing the weight is now:

$$\text{weight} = \frac{1}{\sum_{\text{year=first year}}^{\text{year=last year}} \mu_{\text{year}}} \quad \dots 7$$

$$1 + \frac{\mu_{\text{year}}}{\phi}$$

Note that equations 2 and 7 are identical when all the μ 's are the same. With $v=2.05$ and $\Gamma\mu_{\text{year}}=23.781$, the weight = $1/(1+23.781/2.05)=0.0794$.

Step 3. Estimates.

Now the expected number of accidents for the specific road section at hand and the period 1989-1997 is $0.0794 \times 42.846 + 0.9206 \times 74 = 71.52 \pm (0.9206 \times 71.52) = 8.11$. Note that this estimate is based on the full nine-year accident history and this explains the small weight attached to what is expected at similar sites. The estimate for any specific year is now computed by multiplying the estimate for the entire period by the ratio $\mu_{\text{year}}/\Gamma\mu_{\text{year}}$. Thus, for 1997 the estimate is $[71.52 \pm 8.11] \times 2.710/23.781 = 8.15 \pm 0.92$. These values are listed in row 6. In this manner, the evidence of the entire accident record of nine years is brought to bear on the estimate in any specific year.

Numerical Example 9 - Accounting for secular trend.

In the preceding example the underlying assumption was that while ADT changed over the years, other factors affecting the safety (weather, vehicles, drivers etc.) remained unchanged. However, most everything changes with time. This 'secular trend' can be expressed in multivariate models by 'yearly multipliers' which can be estimated together with all other regression coefficients. Such multipliers are listed in row 2a in Table 4. Thus, e.g., were the model $0.0224 \times \text{ADT}^{0.564}$ applied to data from 1990, it would over-predict the total number of recorded accidents that occurred in 1990 by 1.6%; to bring the prediction and the accident count into agreement one has to multiply by 0.984 as shown in row 2a. The yearly multipliers alter the entries in row 5 and this, in turn, affects all other numerical results.

Table 4

1	Year	1989	90	91	92	93	94	95	96	97	Sums
2a	Yearly Multipliers	1	0.984	1.053	1.005	0.996	0.932	0.931	0.891	0.927	
2b	ADT	4500	4700	5100	5200	5600	5400	5300	5300	5400	
3	Accidents	12	5	9	8	14	8	5	7	6	74
4	μ_{year} [accidents/(km-year)]	2.446	2.466	2.764	2.667	2.756	2.526	2.497	2.390	2.513	23.023
5	Expected accidents in year	4.402	4.439	4.974	4.800	4.960	4.547	4.495	4.301	4.523	41.441
6	Expected annual accident for segment	7.58	7.64	8.56	8.26	8.54	7.83	7.74	7.40	7.79	71.34

Numerical Example 10 - Projection.

The focus so far was on estimating what the expected accident frequency was for some year

August 2001

in the past. Occasionally one wishes to project what accident frequency should be expected at some time in the future. Projections of this kind are always necessary when one wishes compare what safety would have been had some intervention not been implemented to what safety was with the intervention in place. Suppose then that for the segment in numerical example 8 we wish to project the expected number of accidents in 2003 and 2004 when ADTs of 6000 and 6300 are expected and for when the yearly multiplier values of 0.9 and 0.92 are projected.

The starting point for the projection can be any of the values in Table 4. Thus, e.g., the value of 7.79 accidents in 1997 is for $AADT_{1997}=5400$ and the yearly multiplier of 0.927. Recall that the exponent of ADT in the model equation is 0.564. Thus, the projection ratio for 2003 is $(0.9 \times 6000^{0.564}) / (0.927 \times 5400^{0.564}) = 1.030$ and for 2002 it is $(0.92 \times 6300^{0.564}) / (0.927 \times 5400^{0.564}) = 1.083$. Therefore for 2003 we project $7.79 \times 1.030 = 8.02$ accidents and for 2002 we project $7.79 \times 1.083 = 8.44$ accidents.

6. SUMMARY.

The safety of entities is usually estimated from the history of its accident counts. The EB procedure for safety estimation combines accidents counts with knowledge about the safety of similar entities. Doing so has several advantages. Precision of estimation is enhanced when the accident record is sparse and the regression to mean bias is eliminated. As usually, improved precision requires added information. In this case one needs estimates of the Safety Performance Functions for similar entities and an estimate of the applicable overdispersion parameter. Since these are now more widely available, EB estimation of safety should be the preferred practice. The purpose of this paper is illustrate that what may seem to be a complex theory can be put into daily practice.

REFERENCES

1. Hauer, E. (1997). Observational before-after studies in road safety. Pergamon, Oxford, U.K.
2. Datta, T.K., Feber, D., Schattler, K., Datta, S., Effective safety improvements through low cost treatments. Transportation Research Record 1734. Pp. 1-6, 2000.
3. Abbess, C., Jarret, D., and Wright, C.C., Accidents at blackspots: estimating the effectiveness of remedial treatment, with special reference to the "regression-to-the-mean" effect. *Traffic Engineering and Control*, **22**, No. 10, 1981, pp. 535-542.
4. Harwood, D.W., F.M. Council, E. Hauer, W.E. Hughes and A Vogt, Prediction of expected safety performance on rural two-lane highways, Report No. FHWA-RD-99-207, Federal Highway Administration, 2000.
5. Hauer, E., Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation. *Accident Analysis and Prevention*, 33, 799-808, 2001.

August 2001

6. Maycock, G. and Hall, R.D., (1984), Accidents at 4-arm roundabouts. Laboratory Report LR1120, Transport Research Laboratory, Crowthorne, Berkshire, UK.
7. Hauer, E., Ng, J. C. N., and Lovell, J.,(1989), Estimation of safety at signalized intersections. *Transportation Research Record*. **1185**, 48-61.
8. Bonneson, J. A. and McCoy, P. T.,(1993), Estimation of safety at two-way stop-controlled intersections on rural highways. *Transportation Research Record*. **1401**, 83-89.
9. Miaou, S-P.,(1996). Measuring the goodness of fit of accident prediction models. FHWA-RD-96-040. FHWA, McLean, VA.
10. Vogt, A. and Bared, J.,(1998), Accident models for two-lane rural segments and intersections. *Transportation Research Record*. **1635**, 18-29.
11. Persaud, B.N., and Dzbik, L.,(1993) Accident prediction models for freeways. *Transportation Research Record*. **1401**, 55-60.
12. Flowers, R. J.,(1981). A procedure for predicting future accident costs at specified locations. Texas Transportation Institute, Texas A&M University, TARE-53
13. Heydecker, B. G. and Wu, J. (1991) Using the information in road accident records. Proceedings of the 19th PTRC Summer Annual Meeting, Seminar K. PTRC, London.