SHORT-TERM AND AREA-WIDE EVALUATION OF SAFETY MEASURES

Background paper Seminar on Short-term and area-wide evaluation of safety measures, Amsterdam, The Netherlands, 19-21 April 1982, pp. A-F.

R-82-14

S. Oppe & F.C.M. Wegman

Leidschendam, 1982

Institute for Road Safety Research SWOV, The Netherlands

1. INTRODUCTION

In this seminar we will restrict ourselves to safety measures. Thus only countermeasures that are intended to reduce accidents and the consequences of these accidents will be regarded. Of course it is of interest to know how safety measures are related to countermeasures that have been taken for other reasons but also have an impact on safety. Furthermore it is important to know the limitations with regard to goals of higher priority. We can ask ourselves if there are countermeasures on a higher level, such as town planning, traffic distribution, energy consumption etc., that influence safety or the safety measures that have been taken. Or, given these priorities, whether or not the safety measures are optimal with regard to safety or with regard to cost/benefit, cost/effectiveness etc.
We can also investigate the side effects of the safety measures on other subjects such as noise, air pollution and comfort.
However, we propose to restrict ourselves to the evaluation of safety measures with regard to safety.
A second limitation is concerned with the addition of "area wide".
In those cases where traffic circulation has been changed it is always advisable not to restrict the evaluation of countermeasures to the locations where they are implemented. E.g. signalisation of intersections may result in changes in traffic flow and finally in a shift of accidents rather than a reduction. However, here we are concerned with a different thing.

Recently more and more attention has been paid to area-wide traffic plans and proper evaluation studies for these plans. Specific methodological and statistical problems result from this situation. Uncertainties about the effects of new traffic management or environmental schemes result in experimentation and evaluation. The uncertainties arise from the implementation of various different types of countermeasures in "unique" situations.
Short-term evaluations are therefore of interest not only for the policy makers but also for the investigators themselves. A second, rather cynical reason for evaluation is that it is sometimes easy to bypass the public participation process by promising such an evaluation study.

In general, however, the reason behind these studies is to improve an implemented scheme or at least specific countermeasures that are part of it. Therefore, it is necessary to understand why changes do appear. Global studies which result in some percentage of accident reduction do not satisfy the investigator completely. To understand the effects of countermeasures it is inevitable to detect the relevant conditions and to describe quite precisely how these conditions are related to the effects. Furthermore it is important to know how the results of the evaluation studies can be generalised in order to design new schemes.
Finally, two practical problems with regard to area-wide evaluation studies are important. The first one is the definition of the experimental area, the area that is not experimental itself but is influenced by the countermeasures and the control area. The second problem is how to cope with behavioural studies in small streets with low traffic volumes without influencing the traffic process.
Two main reasons for evaluation are mentioned: evaluation of the product of these countermeasures (decrease in unsafety) and evaluation of the changes in the process of traffic behaviour that are initiated with the safety measures. The product is of major concern for policy makers, the process for the investigators, although this distinction is not absolute. The question whether or not these countermeasures are effective, mostly stated in terms of hypothesis testing, is mainly concerned with the product. If explanations are wanted for the presence or absence of effects, mostly stated in terms of parameter estimation, one is mainly concerned with the process. If one is concerned with the product then questions arise such as: "Is this specific conflict measure an acceptable surrogate measure for accidents?". To test such a hypothesis, one investigates whether or not the correlation between the number of accidents and that conflict measure departs from 1. If an investigator regards a conflict to be a behavioural determinant of accidents, then he might ask himself to what extent this aspect of behaviour is related to unsafety, e.g. by asking how much the correlation between this conflict measure and the number of accidents deviates from 0. In this context we may find here all kinds of relational studies, with regard to a large number of behavioural aspects, road conditions and circumstances.
In view of product evaluation, short-term evaluation has the advantage that the influence of disturbing factors is reduced. However, the time

for evaluation may be too short to collect sufficient data, especially when residential areas are investigated. Furthermore, if we want information about the effect of countermeasures on specific aspects of safety, such as pedestrian accidents with children involved, then the problem of insufficient data becomes even more important.

2. PRODUCT EVALUATION

2.1. Safety data

If we want to evaluate the effects of countermeasures, we will measure unsafety before and after the countermeasure and see to what extent there is a reduction in unsafety due to these. If we define unsafety in terms of the probability of accidents and the resulting damage to persons and properties, then the only way to measure unsafety directly, is to count accidents and registrate their effects. We have to estimate probabilities from counts and even with well defined probabilistic models, such as the model of Poisson-distributed accidents, it is very difficult to detect differences in probabilities. In Figure 1, we find an example of the expected number of observations needed to detect various percentages of reduction of accidents. These data result from exact tests applied to Poisson-distributed accidents. A reduction of 20% of the accidents is not significant anymore at a 5%-level, with less than 80 accidents.
A reduction of 30% is only significant with more than 35 accidents. We have to wait for years to get these figures in most situations, both because the kind of countermeasures result in a larger reduction, are very rare and the areas involved have low traffic volumes and therefore low accident numbers (although these numbers may be relatively high as compared to situations with high traffic volumes). There seems to be no way out of this problem that is stated over and over again.
Apart from this fundamental problem there are many methodological problems related to the comparison of accident figures. The report from OECD Road Research Group TS4 (OECD, 1982) gives an excellent survey of problems such as the definition of the correct control group, finding the correct sampling procedure etc. We shall not go into detail on this here. There are also modern techniques that improve the possibilities of comparing data. E.g. De Leeuw & Oppe (1976) describe a log-linear model in which it is possible to compare accident rates of multi-way tables by means of various kinds of hypotheses.
Recently it is even possible to use exact tests to test larger tables than 2x2 designs more efficiently.
However, for each comparison that is stated by means of counts or measures such as accident rates deduced from counts, the problem of

insufficient data remains the most important once the effect of safety measures is evaluated.
The only way to improve statistics is to increase the counts.
One way to do this is to extend sample time. However, with short-term evaluation this seems to be hardly the answer. Furthermore there are difficulties resulting from the increased variation in circumstances that disturb the comparison if time of observation is extended. Another way out is to enlarge the area or increase the number of areas. However, this presumes careful comparisons of situations in order to see if this justifiable. We will return to this in Section 3.
Finally, we can use surrogate measures of safety, e.g. counting conflicts instead of accidents.

2.2. Conflict data

If we want to know how unsafe a particular area is, we really want to establish the accident potential and not just how many accidents have already occurred there. Especially if these areas of concern have very new and drastic designs. Sometimes, we try to estimate the accident potential by means of other indicators of unsafety than the accidents themselves. Sometimes the number of traffic conflicts is used as an indicator of traffic unsafety.
Experimental evidence shows that the "serious-conflict" measure is a better predictor for accidents than the total number of conflicts. Therefore, the definition of conflicts with regard to seriousness is very important in order to improve the validity of the conflict technique. Figure 2 demonstrates an imaginary comparison between two locations. If we choose a definition with regard to the seriousness of conflicts, we select a point on the x-axis. On the surface underneath the curve, right from this point, we find for each location the number of conflicts. If we select the point marked "conflicts" we see that the estimated number of accidents is larger for location A than for location B. If we choose "serious conflicts", then both numbers are more or less equal, while, using the number of accidents, this number if smaller for A than for B. The use of accident rates as a measure of unsafety instead of accidents totals is implicitly based on the assumption that the curves are not parallel. If the curves were decreasing at exactly the same rate for all

locations, the correlation between accidents and accident rates would be perfect. In this case the denominator (some measure of exposure that is used in the accident rate) would give us the best estimation of the potential danger at the locations, because we can measure exposure the most reliably. We know that this is not the case; the accident rate gives us extra information. We have to study the curves in detail before we know if these rates give us sufficient information.

2.3. Accidents, conflicts and exposure

If we look again at Figure 2 we must realise that we use rather far going assumptions in trying to predict the small surface area at the right of the accident point, from the large surface at the right of the conflict or serious conflict point. At least the shape of the curve seems relevant. The accident rate gives us information about this shape. If we define a conflict rate in the same way, using conflicts as some measure of exposure and serious conflicts as a measure of unsafety, then we have some information about the shape of the curve. If well defined, both measures will be more reliable than accident counts, because of the larger number of counts. One problem does limit the relevance of both measures of exposure and unsafety, namely, the problem of the validity of the conflict measures. If we ask for the "content validity" of the conflict measure using conflicts as a measure of exposure we need an operational definition of exposure in each particular case.
Exposure measures deduced from gross traffic -olume data, such as exposure data for pedestrians deduced from time spent in traffic or distance travelled, seem to be insufficient. Especially in situations we are interested in, e.g. residential areas. It seems better to define first situations that are relevant such as the number of encounters between road users, in order to detect which of these situations are critical.
However, it will be even more difficult to find a correct operational definition for serious conflicts, for the detection of critical situations. Content validity seems to be the (very important) first step. Only the "predictive validity" with regard to accidents can inform us about the relevance of the serious-conflict measure as a measure of unsafety. The content validity can inform us only about the "face value"

of the method, or in other words, how relevant the definition looks at first sight.

The face value of the existing conflict measures with regard to situations that are special for residential areas is not high. Many techniques are being developed for dense traffic arterials and/or car-to-car conflicts. It is important to know the relevant cues for accidents between cars and other road users such as pedestrians and bicyclists. Improving the content validity of the conflict measure will be the necessary link between the conflict technique used as a surrogate measure of unsafety and the conflict-analysis technique. This technique regards conflicts as behavioural aspects of the traffic process amongst (and related to) other aspects of behaviour, such as speeds, manoeuvres etc., under various conditions, in order to find explanations for the hazard of specific traffic situations.

3. PROCESS EVALUATION

The effectiveness of countermeasures that have been taken to improve safety, results from the extent to which it is possible to reduce unsafe behaviour or to improve conditions that cause unsafety. If one wants to know if a measure is indeed effective (or as effective as has been assumed), the question arises whether that measure has the intended effect on behaviour or conditions. This answers the question whether or not the measure can have an effect on safety. The next question then is: "Does this change in behaviour or conditions reduce unsafety as was supposed?", or: "Is the measure relevant with regard to safety?"
We can skip the first question and only look at the impact of the countermeasure on safety as has been described as product evaluation. But if such evaluation is not possible, because there are not enough accident data to test these effects, we can ask ourselves whether it is still the best procedure to evaluate the assumed effect on safety by means of surrogate measures instead of evaluating whether or not the countermeasures have the expected impact on traffic behaviour and traffic conditions. Especially in case knowledge about the effectiveness of a particular countermeasure is scarce, it is of great importance to register, apart from the effect on behaviour and conditions that are supposed to be directly influenced in the other conditions and circumstances that existed in that situation and to measure the effect of the countermeasure on the relevant traffic characteristics. We can ask ourselves in what way the characteristics that are supposed to change, are influenced by the countermeasures and which conditions are relevant for this change.
This asks for relational studies in complex situations, especially if the countermeasures that have been taken are compounds of various area-wide countermeasures. For this kind of investigations it is necessary that the results of various situations are collected in order to find relations that are stable and can be generalised for other situations. For example, if we want to evaluate the usefulness of humps in streets, we have to determine under what conditions and in which situations a sufficient reduction in speed will result and how we can cope with dangerous side effects in various situations.
We can possibly find a way of applying a countermeasure which is optimal

for a moderate range of situations, but it is of great importance to know the results outside this range. A large-scale evaluation of countermeasures applied to a diversity of small-scale situations is needed to collect this kind of knowledge.
The registration of conditions is primarily important if we want to compare results from other studies, especially in different countries. Furthermore the registration of discrepancies between the data that is wanted for optimal investigation and the data that were available at the moment of investigation can help to improve future investigations.
If conflict techniques are used to analyse conflicts, in order to get a greater insight into the relation between various countermeasures and conditions on the one hand and the impact of these on the behaviour of road users on the other, this will also result in a better content validity of this technique as a surrogate measure of unsafety.
However, the conflict-analysis technique is only one method of investigation.
All kinds of behavioural measures are available, ranging from sophisticated registration of eye movements and galvanic skin responses to the measuring of velocities or observation of the crossing behaviour of pedestrians.
Because not much is known about the influence of various traffic conditions on the behaviour of road users, the expectations about the effectiveness of measures are based on rather vague theories.
Also little is known about the relation between the estimated risk of situations and feelings of unsafety of road users and the effect of these on their behaviour in traffic or their participation in traffic, vehicle choice or choice of routes. Although these feelings of unsafety may not be direct criteria for the evaluation of safety measures, they are relevant to investigate the relation between the behaviour of road users and safety or the effectiveness of safety measures.
A major problem in the study of this kind of relations is the fact that many characteristics are of a qualitative nature. However, there are recently developed techniques (Gifi, 1981) which can be used to analyse relations between qualitative characteristics as well.

At last, in order to give a complete evaluation of countermeasures and to find an adequate evaluation procedure, the purpose of the measure has to be stated explicitly, together with the means by which one tries to realise this purpose, the expectation about the effectiveness of the measure and its side effects.

REFERENCES

De Leeuw, J. & Oppe, S. (1976). The analysis of contingency tables. Log-linear Poisson models for weighted numbers. R-76-31. SWOV, Voorburg, 1976.

Gifi, A. (1981). Non-linear multivariate analysis. Leyden State University, Department of Data theory. Leiden, 1981.

OECD (Road Research Group TS4) (1982). Methods for evaluating road safety measures. OECD, Paris, 1982.

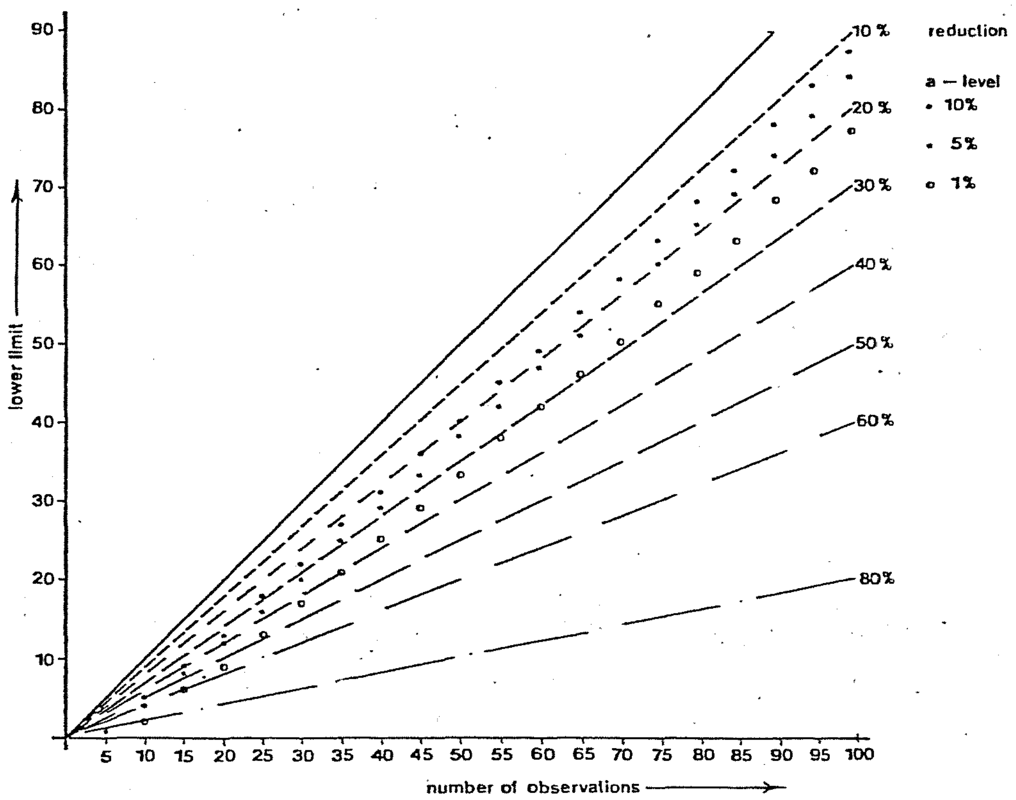

Figure 1

Figure 1 shows the numbers of observations necessary to detect significant reductions in those numbers for values of zero to hundred observations. It is assumed that the number of observations is Poisson distributed. When the number of observations is 100, a reduction of 13% (87 observations) is already indicative (at a 10%-level of significance). A reduction of 16% is already significant at a 5%-level and a reduction of 23% is significant at a 1%-level of significance. A reduction of 20% is not significant anymore at a 5%-level if the number of observations is less than 80. With numbers of observations lower than 50, this reduction is not even indicative.

A reduction of 30% can only be detected at a 1%-level of significance when the number of observations is larger than 60. At a 5%-level with numbers larger than 35. If the number of observations is smaller than 25, a reduction of 30% is not indicative anymore.

A reduction of 40% can only be detected at a 1%-level if the number is greater than 35 and if the level is 5% then only with numbers greater than 20.

If the number of observations is smaller than or equal to 10 even a reduction of 60% is not significant anymore at a 5%-level.

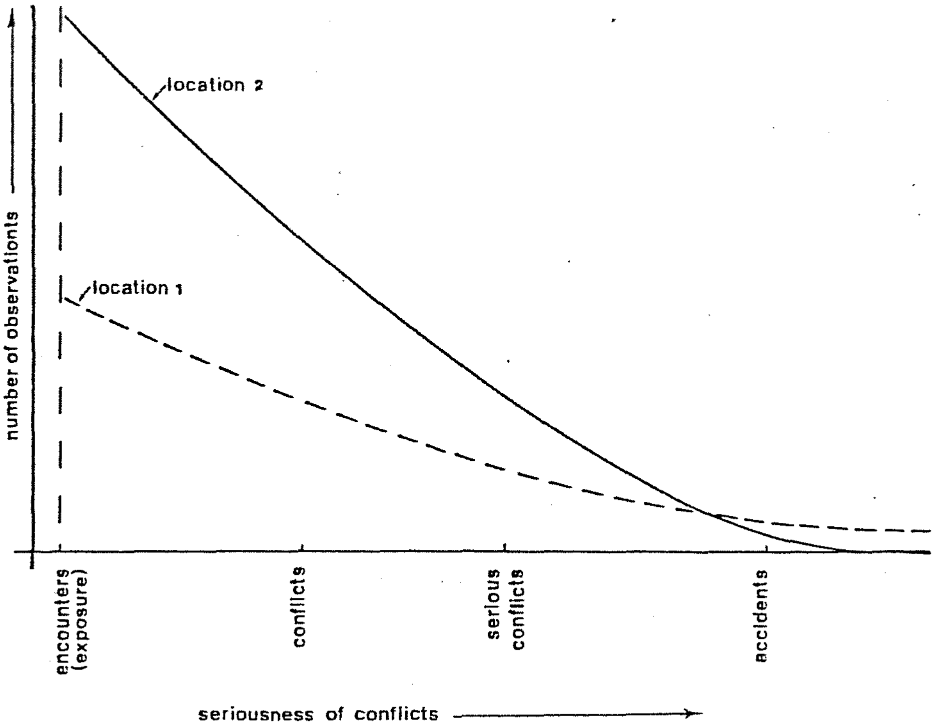
number of observations
location 2
location 1
encounters (exposure)
conflicts
serious conflicts
accidents
seriousness of conflicts

Figure 2