ASSESSING THE VALIDITY OF ROAD SAFETY EVALUATION STUDIES BY ANALYSING CAUSAL CHAINS

Rune Elvik

Institute of Transport Economics
PO Box 6110 Etterstad
N-0602 Oslo, Norway
E-mail: Rune.Elvik@toi.no

INTRODUCTION

Most road safety evaluation studies evaluate the effects of a certain road safety measure on the number of accidents, accident rate or the severity of injuries without describing how and why a certain safety measure produces the observed changes in accident occurrence or injury severity. It has even been argued knowing what “causes” accidents is irrelevant for accident prevention: “An umbrella is equally useful no matter your philosophy about the origin of rain” (Haight 1980, 53). It is true that there is no obvious and simple relationship between the importance of factors that contribute to accidents and the emphasis that ought to be put on trying to change those factors in order to prevent accidents. Human factors, in a wide sense of that term, contribute very much to accidents. However, the most effective way to prevent accidents is often to modify the technical components of the transport system (roads and vehicles), rather than trying to make man infallible (Grime 1987).

It is, on the other hand, obvious that no road safety measure can prevent accidents or reduce injuries unless it affects one or more factors that systematically contribute to accidents or injuries. Let us denote these factors as risk factors. Road safety measures affect risk factors, which in turn affect the incidence of accidents or injuries. The relationship between road safety measures and accidents or injuries, by way of one or more risk factors, will be denoted as a causal chain. This term is intended to suggest that the factors are statistically linked to each other and that they form a logical or temporal sequence, which can be depicted in the form of a causal diagram (Asher 1976). Use of the word “causal” does, however, not imply that a deterministic relationship exists between risk factors and accidents or injuries.

In a paper published thirty years ago, Hall and O’Day (1971) recommended the use of a “causal chain approach” to road safety evaluation studies. In this approach, the direct effects of a road safety measure on one or more risk factors are measured. Effects on risk factors are then related to ultimate outcomes (accidents or injuries), by analysing the chain of events, which follows upon the introduction of a safety measure. To give an example, spreading salt on a road removes snow and ice (or prevents it from forming in the first place) from the road surface. This improves road surface friction, which in turn shortens stopping distance. All else equal, a shorter stopping distance implies fewer, and perhaps less serious accidents. This causal chain can be written as: Spread of salt → Removal of snow and ice from road surface → Improved road surface friction → Shorter stopping distance → Fewer accidents. By analysing causal chains, it is possible to assess the validity of a road safety evaluation study. The objective of this paper is to show
how this can be done by means of two case illustrations. The main research problem addressed in this paper is:

How can one assess the validity of road safety evaluation studies by analysing causal chains, which generate the effects of a road safety measure on accidents or injuries?

**ASSESSING STUDY VALIDITY BY ANALYSING CAUSAL CHAINS**

The validity of a road safety evaluation study denotes the extent to which, there is reason to believe that its results closely approximate the truth (Elvik 1999). If, for example, a study relied on a large and representative sample, employed a rigorous design that controls well for confounding variables, and otherwise has no known flaws, there is strong reason to believe that its results reflect the true effect of the road safety measure it evaluated. If, on the other hand, a study has serious methodological flaws, its results will not necessarily reflect the true effects of a road safety measure.

It is generally believed (Blalock 1961, Hall and O’Day 1971, Asher 1976, Cook and Campbell 1979, Hauer 1997, Elvik 1999), that the validity of a road safety evaluation study is enhanced if that study gives an explicit account of the causal chain of effects, which goes from the road safety measure through one or more risk factors, on to final outcomes, stated in terms of changes in the number of accidents or severity of injuries. If, for example, the speed limit is lowered, we would expect the mean speed of traffic to go down. Lower speed, in turn, leads to fewer and less serious accidents. We tend to place more confidence in a study, which shows this to be the case, than in an otherwise identical study which does not include any information about changes in speed. Accounting for the causal chain through which a measure is effective explains why the measure is effective.

Information about causal chains can be used to assess the validity of road safety evaluation studies by relying on elementary statistical properties of causal chains (Blalock 1961, Asher 1976, Cook and Campbell 1979, Ringdal 1987). Imagine a causal chain $X \rightarrow Y \rightarrow Z$. Then, according to well-established theory, the correlation (Pearson’s $r$) between $X$ and $Z$ is simply the product of the correlations between $X$ and $Y$ and $Y$ and $Z$ (Blalock 1961, 68): $r_{xz} = r_{xy} \cdot r_{yz}$. This means that effects are attenuated along the causal chain. If $r_{xy}$ and $r_{yz}$ both equal 0.6, then $r_{xz}$ will be 0.36. If the correlation between $X$ and $Y$, and $Y$ and $Z$ have the same sign then the correlation between $X$ and $Z$ will be positive. If the correlations (or path coefficients) between $X$ and $Z$, and $Y$ and $Z$ have different signs, the correlation between $X$ and $Z$ will be negative. Suppose the causal chain $X \rightarrow Y \rightarrow Z$ has been observed in two groups, A and B. If $X$ has a larger effect on $Y$ in group A than in group B, and all other statistical relationships are the same, one would expect $X$ to have a larger effect on $Z$ in group A than in group B.

By relying on these basic statistical properties of causal chains, it is possible to define at least three ways in which an analysis of causal chains can be used to assess the validity of a road safety evaluation study:
1. The sign test
If the relationship between X and Y, and between Y and Z have the same sign, this should also apply to the relationship between X and Z.

2. The effect attenuation test
The effect of X on Z should be smaller than the partial effects of X on Y and Y on Z in the causal chain $X \rightarrow Y \rightarrow Z$, provided both partial effects have the same sign.

3. The dose-response test
Assume that a causal chain $X \rightarrow Y \rightarrow Z$ is observed in two groups, A and B. If X has a larger effect on Y, and Y has a larger effect on Z in group A than in group B, X should have a larger effect on Z in group A than in group B.

If a study passes these tests, its validity is enhanced by the fact that a causal chain has been defined and evaluated as part of the study. If, on the other hand, a study fails these tests, its validity is doubtful. If a study fails the causal chain test, one may suspect that its results are attributable to confounding variables that were not controlled in the study. How to apply this framework for assessing study validity will now be shown by means of two case studies.

**CASE 1: SALTING OF ROADS IN NORWAY**

The first case study is a Norwegian study of the effects on accident rate of salting roads in winter (Vaa 1995). It was a case-control study comparing accident rates on salted and non-salted roads during three winter seasons. Salt is spread in order to prevent snow or ice from forming, or from sticking to the road surface. The causal chain for this measure is shown in Figure 1, which shows the causal chain for the case of Children’s Traffic Club as well, to be discussed below.

Figure 1: Causal chain for two road safety measures: salting roads and membership of children’s traffic club

Panel A. Salting of roads

- Spread of salt
- Removal of snow and ice from road surface
- Improved road surface friction
- Shorter stopping distance
- Fewer accidents

Panel B. Children’s Traffic Club

- Membership of traffic club
- Improved knowledge
- Improved behaviour
- Reduced accident rate
The report (Vaa 1995) provides fairly detailed information on the causal chain, which was observed. The main points of this information are summarised in Figure 2. According to Table 4.1 in the report (Vaa 1995, 21), the road surface was bare 85% of the time on salted roads, and 58% of the time on non-salted roads. Salt was therefore effective in removing snow or ice from the road surface. Road surface friction was measured at a sample location. A mean friction coefficient of 0.570 was observed for the salted road, compared to 0.488 for the non-salted road (Vaa 1995, Figure 4.10, 32). By combining information on mean driving speed given in Table 9.10 of the report (Vaa 1995, 92) and a general formula for estimating stopping distance (Ragnøy 1986), mean stopping distance (deceleration distance) was estimated to 41.1 metres for salted roads and 42.9 metres for non-salted roads.

Figure 2: Observed mean values for salted and non-salted roads in Norwegian study. Source: Vaa 1995

<table>
<thead>
<tr>
<th>Stage of causal chain</th>
<th>Values observed</th>
<th>Salted roads</th>
<th>Non-salted roads</th>
<th>Difference in road safety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road surface condition</td>
<td>Bare road</td>
<td>85% of the time</td>
<td>58% of the time</td>
<td>47% gain for salted</td>
</tr>
<tr>
<td></td>
<td>Snow or ice</td>
<td>15% of the time</td>
<td>42% of the time</td>
<td></td>
</tr>
<tr>
<td>Road surface friction</td>
<td>Mean at sampled road</td>
<td>Friction = 0.570</td>
<td>Friction = 0.488</td>
<td>17% gain for salted</td>
</tr>
<tr>
<td>Stopping distance</td>
<td>Metres at mean speed</td>
<td>41.1 metres</td>
<td>42.9 metres</td>
<td>4% gain for salted</td>
</tr>
<tr>
<td>Accident rate (per million vehicle km)</td>
<td>During winter</td>
<td>0.163</td>
<td>0.204</td>
<td>26% gain for salted (adjusted)</td>
</tr>
<tr>
<td></td>
<td>During summer</td>
<td>0.158</td>
<td>0.147</td>
<td></td>
</tr>
</tbody>
</table>

Stopping distance was estimated according to this formula (Ragnøy 1986, equation 5):

\[
S = \frac{v_0^2}{2fg}
\]

in which \(v_0^2\) is initial speed squared (metres per second), \(f\) is the friction coefficient and \(g\) is the gravitational constant (9.8 metres per square second). An initial speed of 77.1 km/h was assumed for salted roads. For non-salted roads, an initial speed of 72.9 km/h was assumed. Effect on accident rate was estimated by taking the odds ratio of winter and summer accident rates for salted and non-salted roads, that is \((0.163/0.204)/(0.158/0.147)\), resulting in a net safety benefit of 26% for salted roads.

At first glance, this analysis would seem to confirm the validity of the estimated safety benefit of salting. All effects observed at the various stages of the causal chain imply a gain in safety for salted roads. The sign of these effects is the same. Moreover, effects are attenuated as one moves from the initial to the final stages of the causal chain. There is, however, an important exception to this conclusion. The adjusted difference in accident rate of 26% seems disproportionately large. It is substantially greater than both the difference in estimated stopping distance, which was only 4% shorter on salted roads, and the difference in road surface friction, for which an improvement of 17% was found for salted roads. On
the other hand, it is difficult to compare the size of effects that were measured differently. One effect was stated as percentage gain (85% bare roads is a gain of 47% compared to 58% bare road), another as a coefficient taking on values between 0 and 1, and a third in metres. Although it is possible to convert different parametric measures of effect size to a common metric (Rosenthal 1994), no way of doing this was found for the different measures used in Figure 2. Hence, the size of effects was compared in percentage terms only.

In principle, it is of course possible to reduce accident rate by 26% even if road surface friction is improved by only 17%, and stopping distance shortened by just 4%. But then, other risk factors would have to be favourably affected. The study gives no evidence that this was in fact the case. On the contrary, one of the most important risk factors, speed, was adversely affected, being higher on salted roads than on non-salted roads. Figure 3 shows retardation curves at mean speed for salted and non-salted roads, assuming a driver reaction time of 2 seconds.
Figure 3: Braking curves for salted and non-salted roads at mean initial speed and mean friction coefficient.

Braking curves for salted and non-salted roads

Initial speed = 77.1 km/h (salt)
Initial speed = 72.9 km/h (no salt)
The curve for salted roads is consistently above that for non-salted roads, suggesting, in effect, that salted roads were less safe than non-salted roads. This is difficult to reconcile with the fact that the mean accident rate in winter, adjusted for the difference in summer accident rate, was 26% lower for salted roads than for non-salted roads. It suggests that part of this difference in accident rate is attributable to confounding variables that were not controlled in the study, rather than the effects of salting.

This suspicion is confirmed by the information given in Table 1. Table 1 compares salted and non-salted roads with respect to potential confounding variables. Traffic volume was substantially greater on salted roads than on non-salted roads. 28% of salted roads had a speed limit of 90 km/h, as opposed to 8% of non-salted roads. This fact alone could explain why mean speed was higher on salted roads than on non-salted roads. Besides, roads with a speed limit of 90 km/h have lower accident rates per million vehicle kilometres of travel than other roads.

Table 1: Comparison of salted and non-salted roads with respect to selected confounding variables. Source: Vaa 1995

<table>
<thead>
<tr>
<th>Variable</th>
<th>Salted roads</th>
<th>Non-salted roads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean AADT (annual average daily traffic)</td>
<td>8,000</td>
<td>2,900</td>
</tr>
<tr>
<td>Percent of road with speed limit below 80 km/h</td>
<td>19%</td>
<td>20%</td>
</tr>
<tr>
<td>Percent of road with speed limit 80 km/h</td>
<td>53%</td>
<td>72%</td>
</tr>
<tr>
<td>Percent of road with speed limit 90 km/h</td>
<td>28%</td>
<td>8%</td>
</tr>
<tr>
<td>Mean speed on dry bare road (sample section)</td>
<td>79.9 km/h</td>
<td>81.3 km/h</td>
</tr>
<tr>
<td>Mean speed on wet bare road (sample section)</td>
<td>77.5 km/h</td>
<td>78.5 km/h</td>
</tr>
<tr>
<td>Mean speed on loose snow (sample section)</td>
<td>68.8 km/h</td>
<td>72.5 km/h</td>
</tr>
<tr>
<td>Mean speed on ice cover (sample section)</td>
<td>70.3 km/h</td>
<td>72.9 km/h</td>
</tr>
</tbody>
</table>

The differences between salted and non-salted roads with respect to traffic volume and speed limit could go a long way towards explaining the observed difference in accident rate. Traffic volume itself affects the proportion of time a road is covered by snow or ice, independently of salting. Traffic, especially cars using studded tyres, wears down snow or ice and ice, leading to a bare road. In the report, two figures are presented, which show the relationship between traffic volume and the percentage of time the road surface was bare for salted and non-salted roads (Vaa 1995, Figures 4.11 and 4.12, 33). A linear function is fitted to the data points in each figure. Figure 4 reproduces these linear functions for salted and non-salted roads.
Figure 4: Relationship between traffic volume and percent of the time the road is bare for salted and non-salted roads. Fitted functions taken from Vaa 1995.

**Relationship between traffic volume and percentage of bare road surface for salted and non-salted roads**

**Mean AADT for salted roads (8000)**

- *Salt*
- *No salt*
It is seen that the percentage of the time the road was bare, is almost independent of traffic volume for salted roads. It should be noted, however, that many salted roads had a much greater traffic volume than the range shown in Figure 4. On non-salted roads, traffic volume has a larger effect on road surface condition. Applying the functions given in Figure 4, it can be estimated that the predicted percentage of time with a bare road at the mean AADT for salted roads is 87% if the roads are salted, and 81% if the roads are not salted. These are the values that ought to be compared to estimate the contribution of salting to improving road surface condition, not the observed mean values as given in Figure 2.

The differences between salted and non-salted roads with respect to speed limit are important for the normal accident rates in the two groups. Normal accident rate for roads with a speed limit of 90 km/h is 0.09 injury accidents per million vehicle kilometres. For roads with a speed limit of 80 km/h, normal accident rate is 0.17 injury accidents per million vehicle kilometres. Finally, for roads with a speed limit below 80 km/h, the normal accident rate can be estimated to 0.35 injury accidents per million vehicle kilometres (Elvik, Mysen and Vaa 1997). Applying these accident rates to salted and non-salted roads, the mean normal accident rates for the whole year can be estimated to 0.182 for salted roads and 0.200 for non-salted roads. This was estimated by applying the road lengths given in Table 1. It is reasonable to assume, however, that the proportion of vehicle kilometres driven on salted roads with a speed limit of 90 km/h is greater than the proportion of road length with this speed limit. Assuming that the distribution of vehicle kilometres of travel on salted roads is 15% on roads with a speed limit below 80 km/h, 50% on roads with a speed limit of 80 km/h, and 35% on roads with a speed limit of 90 km/h, the mean accident rate for salted roads becomes 0.169. This shows that the possibility cannot be ruled out, that the entire difference in accident rate between salted and non-salted roads is attributable to differences in traffic volume, highway design and speed limits, not to the application of salt in winter.

Figure 5 presents an adjusted comparison of salted and non-salted roads. If it is assumed that the friction coefficient is 0.7 on bare roads and 0.25 on roads covered by snow or ice, estimated mean friction is 0.642 for salted roads and 0.615 for non-salted roads. The difference in mean speed between salted and non-salted roads is probably related to the difference in road surface conditions, and does not necessarily mean that drivers increase speed for a given road surface condition. The data given in Table 1 do not suggest that drivers are, in general, going faster on salted roads than on non-salted roads. They are going faster on dry, bare roads than on ice covered roads. The point is that salted roads are more rarely covered by ice than non-salted roads. This leads to a difference in mean speed between salted and non-salted roads for all road surface conditions taken together. Hence, in Figure 5, stopping distance has been estimated assuming an identical initial speed of 75 km/h.
Figure 5: Adjusted values for comparison of safety between salted and non-salted roads in Norway. Derived from Table 1.

<table>
<thead>
<tr>
<th>Stage of causal chain</th>
<th>Estimated values</th>
<th>Salted roads</th>
<th>Non-salted roads</th>
<th>Difference in road safety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road surface condition</td>
<td>Bare road (fitted)</td>
<td>87% of the time</td>
<td>81% of the time</td>
<td>7% gain for salted</td>
</tr>
<tr>
<td></td>
<td>Snow or ice (fitted)</td>
<td>13% of the time</td>
<td>19% of the time</td>
<td></td>
</tr>
<tr>
<td>Road surface friction</td>
<td>Estimated value</td>
<td>Friction = 0.642</td>
<td>Friction = 0.615</td>
<td>4% gain for salted</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(bare = 0.7; snow = 0.25)</td>
<td>(bare = 0.7; snow = 0.25)</td>
<td></td>
</tr>
<tr>
<td>Stopping distance</td>
<td>Estimated value</td>
<td>34.5 metres</td>
<td>36.0 metres</td>
<td>4% gain for salted</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(initial speed 75 km/h)</td>
<td>(initial speed 75 km/h)</td>
<td></td>
</tr>
<tr>
<td>Accident rate</td>
<td>During winter</td>
<td>0.163</td>
<td>0.204</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normal all year</td>
<td>0.182</td>
<td>0.200</td>
<td>12% gain for salted (adjusted)</td>
</tr>
</tbody>
</table>

The adjusted difference in accident rate between salted and non-salted roads is 12%. This is a generous estimate, applying the normal accident rates estimated above, based on road lengths only, and not accounting for possible differences in traffic volume between roads with different speed limits. According to the statistical logic of causal chains, it remains unlikely that the true effect of salting on accident rate could be as great as 12%. It is more likely to be in the order of a 5% reduction in accident rate.

**CASE 2: CHILDREN’S TRAFFIC CLUB IN NORWAY**

In 1966, a traffic club for children was created in Norway. The Club is intended for children between 3 and 6 years of age. The effects on safety of being a member of this club were evaluated in 1974 (Schioldborg 1974). The evaluation was designed as a case-control study, in which members were compared to non-members with respect to accident rate and a number of factors that were believed to affect accident rate. The study was controversial at the time of its publication (Knudsen 1975A, 1975B, Schioldborg 1975A, 1975B). It was criticised for not controlling for potential self-selection bias in choosing to join the club. Those who join, it was argued, differ systematically from those who do not join in terms of variables that affect children’s safety in traffic. In defence of the study, Schioldborg (1975A, 1975B) argued that the results were credible because they could be accounted for in terms of the causal chain by which the club influenced children’s road accident rate. The causal chain modelled by Schioldborg is shown in Figure 1.

Figure 6 shows the mean values observed for boys and girls for knowledge, behaviour and accident rate. For each sex, members are compared to non-members. A knowledge score was formed by assigning points for correct answers. The more points, the better the
knowledge. This comparison was favourable for girls who were members of the club. For boys, no difference between members and non-members was found.

Figure 6: Observed mean values for members and non-members of Children’s Traffic Club in Norway. Source: Schioldborg 1974

<table>
<thead>
<tr>
<th>Stage of causal chain</th>
<th>Observed values</th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Members</td>
<td>Non-members</td>
</tr>
<tr>
<td>Knowledge score</td>
<td>Points scored</td>
<td>21.40</td>
<td>20.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.50</td>
<td>15.75</td>
</tr>
<tr>
<td>Behaviour score</td>
<td>Points scored</td>
<td>61</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>Accident rate</td>
<td>Per 10,000 children</td>
<td>19.4</td>
<td>22.1</td>
</tr>
</tbody>
</table>

Gain in road safety for members

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>2% gain</td>
<td>18% gain</td>
</tr>
<tr>
<td>Behaviour</td>
<td>2% loss</td>
<td>24% gain</td>
</tr>
<tr>
<td>Accident</td>
<td>3% loss</td>
<td>8% loss</td>
</tr>
</tbody>
</table>

Behaviour was observed in real traffic when the child was accompanied by an adult, as well as when it was walking alone. Correct behaviour was noted; the more points, the more often correct behaviour was observed. This comparison favoured boys who were members of the club. For girls, no difference between members and non-members was found. Accident rate, defined as accidents per 10,000 children was compared for members and non-members. Members had a lower accident rate than non-members both among boys and girls. The percentage difference in accident rate was largest for girls.

Do these results lend support to the claim that the observed differences in accident rate were primarily attributable to membership in the traffic club? Applying the elementary logic of causal chains hardly supports this interpretation. The findings fail the sign test, both for boys and girls. For boys, knowledge did not improve, whereas behaviour did, and members had a lower accident rate than non-members. For girls, knowledge did improve, but behaviour did not, but still members had a lower accident rate than non-members. If accident rate is reduced by correct behaviour, the results for boys make sense, whereas those for girls do not. The causal chain fails the effect attenuation test as well. For girls, for example, a large effect is found on knowledge, hardly any effect on behaviour, but a large effect on accident rate. Finally, the causal chain fails the dose-response test. There is a large difference in accident rate for girls, but the differences between members and non-members observed at earlier stages of the causal chain are not consistently larger for girls than for boys. In short, the causal chain leaves an untidy picture, which does not lend strong support to interpreting the observed differences in accident rates between members and non-members as an effect of membership in Children’s Traffic Club.

Once more, the suspicion is aroused that uncontrolled confounding factors may account for the results. The first potentially relevant factor that comes to mind is traffic exposure. No precise data were obtained on this, but children were asked how often they were exposed to traffic. Answers were coded as “often”, “sometimes”, and “rarely” (Schioldborg 1974, Table 7, 31). Let these answers, arbitrarily be translated to once per day, once per week and once
per month, respectively. Mean exposure for boys was then 14% lower among members than among non-members. This difference fully accounts for the observed difference between members and non-members in the number of accidents per 10,000 boys. Mean exposure among girls was 29% higher among members than among non-members. This cannot account for the difference in the number of accidents per child between members and non-members, which had the opposite sign of the difference in the amount of exposure.

Parents of children who were members of the club had a higher mean income and a higher level of education than parents of children who not members of the club. This suggests that their social status was higher. At the time of the study, nothing was known about the relationship between social status and road accidents for children in Norway. In general, however, social status is a powerful predictor of the incidence of a number of problems, including ill health, unemployment, exposure to environmental problems, and, possibly, accidents.

**DISCUSSION**

Ideally speaking, we would not just like to know how effective a road safety measure is in improving road safety. We would also like to know why it is effective – which are the risk factors it influences. Road safety evaluation studies that obtain data on the causal chain which generates the effects of a road safety measure on accidents or injuries are therefore more valuable, and possibly more valid, than otherwise identical studies that leave this causal chain a black box.

As shown in the examples presented in this paper, it is by no means obvious that the results of studies that specify a causal chain of the form: Measure → Risk factor → Accidents, are more valid than the results of studies that do not specify such a causal chain. A causal chain that, on a superficial reading, looks plausible and seems to add credibility to the results of a study, may, on closer inspection, indicate that the results are misleading. The case of salting of roads eminently illustrates this. The observed effect of salting roads on accident rate was implausibly large, considering the size of the effects found at earlier stages of the causal chain. Although all the partial effects observed throughout the causal chain modelled in this study were consistently in the same direction, their magnitudes did not match very well. A closer examination of the two categories of road that were compared found that they differed systematically with respect to a number of potentially confounding factors, which could in principle explain the entire difference in accident rate between salted and non-salted roads.

A similar examination of the causal chain modelled in the study of Children’s Traffic Club in Norway revealed several anomalies in the partial effects observed in the causal chain. These partial effects were untidy, and were not consistent with respect either to direction, magnitude or a dose-response pattern. The causal chain as observed in the evaluation of Children’s Traffic Club hardly supports the claim that the differences in accident rate observed between members and non-members were primarily attributable to membership of the club.

There are three main lessons to be learnt from this. The first lesson is that trying to model the causal chain, which produces the effects of a road safety measure on accidents or injuries, is a very useful part of the design of any evaluation study. If it is not possible to identify, and preferably measure, the risk factors that are supposed to be favourably influenced by a road safety measure, there is simply no reason to believe that the measure will be effective. To be effective, any road safety measure has to influence at least one factor that systematically contributes to accident occurrence or injury severity.
The second lesson is that obtaining data on the causal chain allows for a more systematic – and, by the same token, possibly a more critical – assessment of the validity of study findings than if the causal chain is left in the dark. If the causal chain obeys the elementary statistical properties of recursive causal chains, it adds to the validity of study findings. If, on the other hand, it violates these statistical regularities, it casts doubt on the validity of study findings. Violation of the expected pattern of findings in a specific causal chain suggests that the true causal process that produced changes in the count of accidents or the severity of injuries may have been an entirely different one.

This leads to the third lesson of the case studies discussed in this paper, which is that the most serious threat to the validity of road safety evaluation research is lack of control of confounding factors. This point has been made many times before (see, for example, Elvik 1999), but its importance can hardly be overstated. Not even the most meticulous modelling of causal chains can substitute for a rigorous control of confounding factors. The validity of road safety evaluation studies is affected more strongly by the adequacy of control of confounding factors than by any other single known source of bias in such studies.

CONCLUSIONS

The main conclusions to be drawn from the research reported in this paper can be summarised as follows:

1. A road safety measure affects the number of accidents, or the severity of injuries, by way of a causal chain, which consists of one or more risk factors that the measure is intended to affect favourably.

2. A good road safety evaluation study, explicitly models and collects data about the causal chain by which the measure is hypothesised to be effective. The validity of study findings can then be assessed by analysing the causal chain, in order to determine if findings are consistent with the basic pattern to be expected in any recursive causal chain.

3. Two case illustrations of this type of analysis are presented in the paper. Both are based on road safety evaluation studies reported in Norway. The first case concerned salting of roads. It was found that the size of the effect on accident rate attributed to salting was inconsistent with the size of the partial effects observed at prior stages of the causal chain. On closer examination, the conclusion was reached that study findings were invalidated by an inadequate control of confounding factors. The second case concerned Children's Traffic Club in Norway. For this study, it was found that the partial effects observed at various stages of the causal chain were messy, and did not form any pattern that supported the conclusion that being a member of the club reduced accident rate.

4. The main conclusion from both case studies is that the most important threat to the validity of road safety evaluation studies is lack of control of confounding variables. This threat to study validity cannot be compensated for by analysing the causal chain by which a measure is thought to affect accidents or injuries.
REFERENCES


Knudsen, K. Replikk. Tidsskrift for samfunnsforskning, 16, 259-261, 1975B.


Schioldborg, P. Svarreplikk. Tidsskrift for samfunnsforskning, 16, 263-264, 1975B.
Vaa, T. Salting og trafikksikkerhet. Del 2: Sammenligning av ulykkesfrekvens påsaltet og
usaltet vegnett. Saltingens effekt på kjørefart. Rapport STF63 A95004. Trondheim, SINTEF

**LIST OF FIGURES AND TABLES**

Figure 1: Causal chain for two road safety measures: salting roads and membership of
children’s traffic club

Figure 2: Observed mean values for salted and non-salted roads in Norwegian study. Source:
Vaa 1995

Figure 3: Braking curves for salted and non-salted roads at mean initial speed and mean
friction coefficient

Figure 4: Relationship between traffic volume and percent of the time the road is bare for
salted and non-salted roads. Fitted functions taken from Vaa 1995

Figure 5: Adjusted values for comparison of safety between salted and non-salted roads in
Norway. Derived from Table 1

Figure 6: Observed mean values for members and non-members of Children’s Traffic Club in
Norway. Source: Schioldborg 1974

Table 1: Comparison of salted and non-salted roads with respect to selected confounding
variables. Source: Vaa 1995