THE IMPORTANCE OF CONFOUNDING IN OBSERVATIONAL BEFORE-AND-AFTER STUDIES OF ROAD SAFETY MEASURES

Rune Elvik

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

INTRODUCTION

Nearly all studies that evaluate the effects of road safety measures are observational, that is non-experimental studies in which treatment is not assigned at random. One of the most common study designs employed in observational evaluation studies of road safety measures, is a before-and-after study with a comparison group. The most serious threat to the validity of the results of studies employing this design is lack of control of potentially confounding variables. Depending on how the study is carried out, observational before-and-after studies may control for a number of important confounding variables. It is, however, always possible to argue that the control of confounding variables in observational before-and-after studies is imperfect.

In experimental studies, the random assignment of subjects to the treatment and control conditions ensures, at least in large samples, that all confounding variables are controlled. There is no need to identify in advance the confounding variables that need to be controlled. In observational studies, on the other hand, this has to be done. Unfortunately, there is no strong theoretical basis for identifying potentially confounding variables in observational road safety evaluation studies. Hence, the variables that are regarded as confounding tend to be those that have been found to confound the results of past studies. Moreover, lack of relevant data imposes limits on the confounding variables that can be controlled. It is hardly surprising that controversies about the interpretation of observational road safety evaluation studies often centre on the control of confounding variables in these studies. A few examples of such controversies are presented in this paper.

This paper discusses a number of issues that concern the control of confounding variables in observational before-and-after studies of road safety measures. The most important questions to be discussed are:

1. What is confounding?
2. What are the most important confounding variables in observational before-and-after studies of road safety measures?
3. How can various confounding variables best be controlled in observational before-and-after studies of road safety measures?
4. To what extent do commonly recognised confounding variables affect the results of before-and-after studies? Would lack of control of these variables serious bias the results of observational before-and-after studies?
THE CONCEPT OF CONFOUNDING AND APPROACHES TO CONTROLLING FOR IT

General discussions of the concept of confounding can be found in textbooks in epidemiology (e.g. Kleibaum, Kupper and Morgenstern 1982). In observational before-and-after studies of road safety measures, a confounding factor is:

A. Any factor, other than the measure whose effects the study is designed to evaluate,

B. Which is itself not influenced by the measure (i.e. is exogeneous in a causal diagram), and

C. Which produces changes in the expected number of accidents from the period before to the period after the measure is introduced.

Briefly stated, a confounding variable is any variable affecting the number of accidents or injuries whose effects, if not estimated, can be mixed up with effects of the measure being evaluated. Variables that are commonly regarded as potentially confounding in observational before-and-after studies of road safety measures include (Hauer 1997):

- Regression-to-the-mean
- Long term trends affecting the number of accidents or injured road users
- General changes of the number of accidents from before to after the road safety measure is introduced
- Changes in traffic volume
- Any other specific events introduced at the same time as the road safety measure

Various approaches can be taken to control these variables. The textbook by Ezra Hauer (1997) provides an excellent survey of these approaches. There are two main approaches: (1) Estimate the effects of a confounding factor statistically, and (2) Use a comparison group. In many cases, both approaches will be used in the same study. A comparison group is generally assumed to control for all confounding factors whose effects cannot be estimated directly. It is not always obvious which confounding factors that are controlled by using a comparison group, and which have to be estimated statistically. An example is given below of a study in which it seems reasonable to assume that the comparison group controlled adequately for changes in traffic volume, whereas the author of the study assumed that this confounding factor had to be controlled by statistical estimation (Sæverás 1998).

Before proceeding, the difference between long term trends and general changes of the number of accidents should be explained. A long-term trend is a systematic tendency for the number of accidents to increase or go down during a period of several years. Figure 1 shows an example of long-term trends towards fewer road accident fatalities in the Netherlands and Belgium in the years before the Netherlands adopted a quantified road safety target in 1986 for the year 2000. These trends are very similar in the two countries, which is why Belgium was regarded as a suitable comparison country for the Netherlands.
Figure 1: Long term trends in road accident fatalities in Belgium and the Netherlands before and after the Netherlands set a quantified road safety target.

General changes of the number of accidents denote changes from the period before the measure to the period after, when all years before and all years after have been combined. Figure 2 shows such changes for the Netherlands and Belgium, based on the before and after periods in Figure 1. It is rather common in observational before-and-after studies of road safety measures to pool data for all years before and all years after the way it has been done in Figure 2. This results in a loss of information whenever a long-term trend is discernible. In many cases, however, finding a long-term trend will be difficult, because accident data are available for too few years or subject to large random fluctuations from year to year.
Figure 2: Changes in the total number of road accident fatalities in Belgium and the Netherlands from before to after the Netherlands set a quantified road safety target, combining all years before and all years after

How to control adequately for long-term trends is a complex issue and will not be further discussed in this paper.

**STUDY APPROACH**

The subsequent analysis can be divided into three sections:

1) A presentation of some controversies about the interpretation of studies in which the results were greatly affected by the choices made by researchers regarding control of confounding factors.

2) A discussion of how best to control for various confounding factors, in particular changes in traffic volume, in observational before-and-after studies of road safety measures.

3) Case studies of the size of the effects of two confounding factors in observational before-and-after studies of road safety measures, and a discussion of the importance of controlling for these factors based on the case studies.

The discussions about control of confounding factors took place in various journals (Payne and Barmack 1964, Imhoff 1964, Harvey and Durbin 1986, Ehrenberg 1986, Elvik 1990A, Strand 1990). The study serving as the basis for discussing how to control for changes in traffic volume was conducted by a regional highway agency in Norway and reported in 1998 (Sæverås 1998). Norwegian studies that have evaluated seven different road safety measures were chosen to give examples of the effects of confounding factors. The studies evaluated the effects of:

A. Traffic separation, in the form of tracks for walking and cycling, physically separated for motor traffic (Elvik 1990B, Dietrichs 1991, Thingwall 1991),
B. Bypass roads around 20 small towns in Norway (Amundsen and Hofset 2000; Elvik, Amundsen and Hofset 2000)

C. New urban arterial roads in the towns of Bergen (Sæverås 1998), Oslo (Amundsen and Gabestad 1991) and Trondheim (Holt 1993),

D. Widening the road from two to four travel lanes and adding a median barrier on an urban arterial road in the town of Trondheim (Langeland 1999),

E. Black spot treatment, mostly by means of junction improvement (Elvik 1985, Christensen 1988),


G. Speed cameras (Elvik 1997)

All the studies that evaluated the effects of these road safety measures were made in the same country and refer to effects on injury accidents. The studies employed identical designs that controlled for regression-to-the-mean, by means of the empirical Bayes method (Hauer 1997), and general changes of the number of accidents, by means of a comparison group. The results of studies that evaluated the same road safety measure, were combined by applying the log odds method of meta-analysis (Elvik 1999). The studies of Eick and Vikane (1992), Eriksen (1993) and Stigre (1993) were re-analysed to control for regression-to-the-mean by relying on the study of Sakshaug (1998). The study of Sæverås (1998) was also re-analysed, see the discussion below.

CONTROVERSIES ABOUT CONTROLLING FOR CONFOUNDING IN ROAD SAFETY EVALUATION STUDIES

To researchers, the need to control for known confounding factors seems obvious. However, not everybody shares this point of view, as evidenced by some historical controversies about controlling for confounding factors in road safety evaluation studies.

Chris Imhoff (1964) launched the following attack on an experimental evaluation of the Smith-Cummings-Sherman driver training system, performed by Payne and Barmack (1964):

“There seems to be a trend in accident research today to shoot down the old and honourable concepts that have been the mainstay of successful fleet accident prevention programs for many years. ... It was reported that questionnaires had been sent out to a list of 49 fleets who had used the Smith system. The list, incidentally, was supplied by Harold Smith himself. The replies to this questionnaire indicated that there were small improvements in the accident rates of most of the fleets that had used the Smith system. Now this finding alone should have been sufficient to prompt the abandonment of this research project and moving on to some more important problem.”

Payne and Barmack replied:

“If progress is to be made in controlling and preventing accidents, we know of no alternative to rigorous, objective testing of the programs offered to reduce accidents. All of us who are interested in the problem should be prepared to accept our disappointments, and to learn from them.”

Ehrenberg (1986) criticised an elaborate statistical model developed by Harvey and Durbin (1986) in order to evaluate the effects of seat belt legislation in the United Kingdom.

“This really is a dreadfully vapid paper – full only of intellectual masturbation. The danger is that younger statisticians often think such performances sophisticated and ape them. Why did the authors not simply report that all their “time-series” mumbo-jumbo made no
difference? ... Anyone with enough courage like Dr Chatfield to do simple percentages on the raw data as in Table 1 would have “found” much the same.”

The reply from Harvey and Durbin was this:

“It is difficult to respond to professor Ehrenberg’s tirade since it contains little in the way of relevant rational argument. Furthermore the points he does make seem naive. For example, he claims that a simple before and after comparison gives virtually the same results as our analysis. ... The fact that a simple before and after comparison sometimes gives the same result as a model which allows for a trend and explanatory variables does not mean it always gives the same result.”

The third illustration concerns a study that evaluated the effects of tracks for walking and cycling in Norway. Professor Arvid Strand (1990) criticised this study in the following terms:

“Why not present the study with the headline: “Tracks for walking and cycling do not always lead to more accidents.” Or simply: “Unsuccessful attempt to find the effects of cycle tracks on accidents”. The latter headline would, in my opinion, be closest to the truth. ... A simple way of testing the effects of cycle tracks is to compute the number of accidents in the after period as a percentage of the number of accidents in the before period. But Elvik did not have the courage to do this simple comparison. In stead, he constructs complicated models based on data that may not be sufficient to determine the effects of cycle tracks on accidents.”

Elvik (1990A) replied the following to this criticism:

“Yes, Strand is correct that the number of accidents went down. Then why did I not conclude that cycle tracks were effective in reducing accidents? Because there is another possible explanation for this decline, namely regression-to-the-mean. By controlling for regression-to-the-mean, it turns out that this entirely explains the decline in accidents. No effect of the cycle tracks could then be found.”

These three cases have a lot in common. In general, the critics are making three points:

1) It is not necessary to control for confounding factors, because they do not confound the results. Ehrenberg (1986), in particular, makes this point.

2) The wrong method is used to control for confounding - poor data are subjected to sophisticated analyses. Strand (1990) emphasises this point.

3) Controlling for confounding factors destroys the results of the study, because the road safety measure no longer has an effect after confounding factors have been controlled. This is the core of the statement made by Imhoff (1964).

The third of these points cannot be taken seriously. Research methods cannot be evaluated in terms of whether or not we like the results they produce. The other two points, however, deserve to be taken seriously. If a potentially confounding factor does not really confound a study, there would seem to be no need to control for it. But, there is a snag here, as pointed out by Harvey and Durbin (1986). The snag is that it is impossible to know if a potentially confounding factor really confounds or not unless we estimate the effects of that factor, which of course means that we have to control for it.

There are obviously many methods that can be used to control for confounding factors, and some of these methods may be less suited than others. The point made by Strand (1990) is, however, misplaced. There is nothing wrong about subjecting accident data to sophisticated statistical analyses, despite the fact that these data are known to be imperfect. On the contrary, resorting to quite complicated analyses is sometimes necessary in order to model the often quite complicated process that produced the data.
THE ROLES OF STATISTICAL ESTIMATION AND COMPARISON GROUPS IN CONTROLLING FOR CONFOUNDING

Sæverås (1998) reports a study that evaluated the effects on road safety of a new arterial road west of Bergen, Norway. The study was a before-and-after study with a comparison group. The study controlled explicitly for:

- Regression-to-the-mean
- Changes in traffic volume
- General changes of the number of accidents in Hordaland county (where the town of Bergen is located)

The before and after periods were both four years. Accident counts for all years before and all years after were combined. There were 323 accidents before and 268 accidents after in the study area. The county of Hordaland was used as comparison area. Table 1 presents the results of the study.

Table 1: Original analysis and re-analysis of a before-and-after studies of the effects on accidents on a new urban arterial road in Bergen, Norway

<table>
<thead>
<tr>
<th>Steps of analysis</th>
<th>Original analysis</th>
<th>Re-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of accidents before</td>
<td>323</td>
<td>323</td>
</tr>
<tr>
<td>Expected after – controlling for regression-to-the-mean</td>
<td>318</td>
<td>317</td>
</tr>
<tr>
<td>Expected after – controlling for traffic growth (and regression to the mean)</td>
<td>372</td>
<td>----</td>
</tr>
<tr>
<td>Expected after – controlling for general changes in the number of accidents (and the other confounding factors)</td>
<td>360</td>
<td>293</td>
</tr>
<tr>
<td>Number of accidents after</td>
<td>268</td>
<td>268</td>
</tr>
<tr>
<td>Estimated safety effect</td>
<td>-25.5%</td>
<td>-8.5%</td>
</tr>
</tbody>
</table>

The roads included in the study had a slightly higher than normal accident rate in the before period. Controlling for regression to the mean, by means of the empirical Bayes method proposed by Hauer (1997) gave an expected number of accidents of 318 for the after period. The report states that traffic volume was 17% higher in the after period than in the before period. To control for this the expected number of accidents (318) was multiplied by 1.17, resulting in an expected number of 372 for the after period. There was a slight decline in the number of accidents in Hordaland at large from before to after. Controlling for this, the expected number of accidents in the after period was estimated to 360. The recorded number was 268, and the effect of the new arterial road was estimated to 25.5% accident reduction.

The most questionable aspect of this analysis is the way traffic growth was controlled in the study. In the first place, the number of accidents was assumed to increase by the same percent as traffic volume. This assumption is not supported by empirical evidence. Traffic grew by 17% in Hordaland county but the number of accidents did not grow by 17%. In general, when traffic grows by X%, the number of accidents tends to increase by less than X%, often considerably less in time-series data sets (Fridstrøm 1999). In the second place, it can be argued that the comparison group captures the effects on accidents of traffic growth. Changes in the number of accidents in the comparison group from before to after reflect the combined effects of all factors that affected the number of accidents, including traffic growth. There is therefore no need to control explicitly for traffic growth the way it was done in this
study. Doing so is a form of double counting, which in this case artificially inflated the expected number of accidents in the after period.

A re-analysis of the study, in which it was assumed that the comparison group reflects the combined effects of all factors influencing the expected number of accidents, is presented in Table 1. In this analysis, all of Norway was used as a comparison group, not just the county of Hordaland. According to this analysis, the expected number of accidents in the after period was 293. The effect of the new urban arterial road was estimated to 8.5% accident reduction. In general, it is reasonable to assume that a large comparison group, i.e. one in which the annual count of accidents is at least several hundred, captures the effects of all factors that may produce changes over time in the long-term expected number of accidents.

THE IMPORTANCE OF CONFOUNDING: SEVEN TEST CASES

Are the effects of confounding factors in general important for the results of observational before-and-after studies of road safety measures? To answer this question, seven test cases were examined. In all these cases, the contributions to the changes of the number of accidents from before to after introduction of a road safety measure were estimated for:

- Regression to the mean,
- General changes of the number of accidents,
- The road safety measures, once regression to the mean and general changes had been controlled.

Figure 3 presents the results. Regression to the mean was estimated to 20-30% in three of the cases, but was of a negligible magnitude in the other four cases. General changes of the number of accidents were in most cases less than plus or minus 10%.

*Figure 3: Percentage change of the number of accident attributed to regression-to-the-mean, general changes and the road safety measure for seven road safety measures*
The effects attributed to the road safety measures once regression-to-the-mean and general changes of the number of accidents had been controlled range from 6% to 52% reduction of the number of accidents. Whenever the effects of a confounding factor are in the same direction as the effects of a road safety measures, not controlling for the confounding factor will lead to an overestimate of the effects of the road safety measure. This is true in five of the seven cases that have been examined.

Figure 4 presents the effects that would be attributed to the road safety measures in simple before-and-after studies, not controlling for any confounding factors, and in the studies that controlled for regression to the mean and general changes of the number of accidents.

**Figure 4: A comparison of controlled and uncontrolled estimates of the effects of seven road safety measures**

On the average, the uncontrolled estimate of effect is a 32% reduction of the number of accidents. The mean controlled estimate of effect is a 20% reduction of the number of accidents. The largest differences between the uncontrolled and controlled estimates of effect refer to traffic separation and black spot treatment. In both of these cases, regression to the mean was found to have a major impact on the results of the before-and-after studies.

**DISCUSSION**

Very many factors influence the number and severity of road accidents. Some of these factors are better known than are others. The most important potentially confounding factors in observational before-and-after studies of road safety measures are fairly well known and can to some extent be controlled. Control of confounding factors must rely on a mixture of an explicit estimation of the effects of confounding factors and an implicit control of them by means of a comparison group.

It is not always perfectly clear exactly which are the confounding factors a comparison group controls for, and which are the factors that must be controlled statistically. This could be a source of controversy regarding the interpretation of the findings of studies that have...
employed a comparison group. The factors that are controlled for by a comparison group depend on how it is chosen. The following general guidelines can be given in this respect:

1) A comparison chosen to match the treatment group with respect to an abnormal accident rate in the before period controls for regression to the mean, but not for any other confounding factors.

2) A general comparison group, comprising a larger area in which the treated sites are located, will normally control for general changes of the number of accidents, including area-wide changes in traffic volume, but usually not for regression to the mean or long term trends in the number of accidents.

3) A general comparison group, comprising a large area and many years of before data, will normally control for long term trends in the number of accidents and all factors that produce such trends, including changes in traffic volume.

These general points of view concerning what a comparison group controls for are summarised in Table 2. A typology of comparison groups used in observational before-and-after studies has been developed on the basis of three criteria:

A. Matching: A matched comparison group is chosen to be as similar to the treatment group as possible with respect to accident experience.

B. Size: A large comparison group represents a considerably larger accident sample than the treatment group.

C. Length of before period: A distinction is made between cases using just one period before and just period after, each of which may comprise several years, and cases using several periods, at least before.

Table 2: A typology of comparison groups used in observational before-and-after studies of road safety measures and the confounding factors normally controlled by each type of comparison group

<table>
<thead>
<tr>
<th>Characteristics of comparison group</th>
<th>Confounding factors normally controlled by using a comparison group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td></td>
</tr>
<tr>
<td>Matched</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Table indicates whether or not a comparison group will normally control for the confounding factor listed. It is seen that comparison groups cannot, in general, control for all the important confounding factors in observational before-and-after studies of road safety measures. Perhaps the most difficult factor to control adequately is long-term trends in the number of accidents. A comparison group can only be assumed to control for long term
trends if it can be shown that the long term trends in the before period were similar in the treatment group and the comparison group (Hauer 1997).

CONCLUSIONS

The main conclusions of the research presented in this paper can be summarised in the following points:

1) The most difficult part of observational before-and-after studies of road safety measures is to obtain an adequate control of confounding factors. It is not always obvious how best to control confounding factors. This may lead to controversy regarding the interpretation of the results of before-and-after studies.

2) Using a large comparison group will in most cases control for a number of important potentially confounding factors, including changes of traffic volume. The effects on accidents of changes in traffic volume will be reflected in the accident counts for a large comparison group.

3) Some objections that have been made to studies that have tried to control for confounding factors were reviewed. Most of these objections are not convincing. The fact that a potentially confounding factors may turn out not to be confounding cannot be known unless an attempt is made to control for that factor.

4) Seven case studies that estimated the effects of two important confounding factors in before-and-after studies were presented. The two confounding factors considered were regression to the mean and general changes of the number of accidents in a larger area. It was found that both these factors sometimes have a major impact on the results of before and after studies. There is a tendency for the effects of road safety measures to be overestimated when these confounding factors are not controlled.
REFERENCES


