27th Australasian Transport Research Forum, Adelaide, 29 September – 1 October 2004

Paper title: **Evidence-based road safety policy? Evidence-based transport policy? A discussion of randomised experimentation and meta-analysis in the evaluation of interventions**

Author(s) name(s): T P Hutchinson and A J Meier

Organisation(s): Centre for Automotive Safety Research, University of Adelaide

Contact details:

Postal address: Centre for Automotive Safety Research
University of Adelaide
South Australia 5005

Telephone: (08) 8303 6086
Facsimile: (08) 8232 4995
email: paul@casr.adelaide.edu.au

Abstract (200 words):
This paper discusses methods of evaluating interventions. Examples in the transport field might include improving compliance with speed limits, or boosting the use of public transport. In recent years there has been a push towards higher methodological standards in medical research, because of the biases that can easily creep into the comparisons that are of central interest in a research project. Particular features of this trend are (a) in the conduct of research, randomised allocation of the experimental units (e.g., people) to treatment or control groups, and (b) in synthesis of previous research, carrying out what is termed a meta-analysis rather than a review in narrative style, with the analysis giving much greater weight to studies that had good methodology than to those which did not. The phrase 'evidence-based' is used in this paper to refer to these features, not to a general hope that research and facts will influence policy. This paper documents the extent to which high methodological standards have been adopted in transport and transport safety to date, and attempts to assess what place they might have in future.

Introduction

It is now over ten years since the establishment in the United Kingdom of the Cochrane Collaboration (see http://www.cochrane.org), a centre for the promotion of evidence-based medicine, primarily through the commissioning and distribution of systematic reviews of health-related interventions. Since then, similar methods have been used, and similar organisations have emerged, in fields such as education, criminology and social welfare. This paper examines the impact of this trend in transport and transport safety research.

For this paper, the phrase ‘evidence-based’ is used, as in medical research, to refer to the employment of randomised treatment and control groups, and the systematic reviewing of such research using meta-analysis. The phrase can be diluted to be hardly more than lip-service to the desirability of having some input to policy from objective research data, but we do not use it in diluted form. (There are other components of evidence-based medicine, notably a commitment to understanding both the benefits and limitations of clinical research, but randomised experimentation and meta-analysis are the most distinctive.)

The aims of this paper are as follows.

- Describe the key features of methods that lead to use of the term ‘evidence-based’.
- Demonstrate that these methods are now being used in some fields of social science where experimentation is really difficult.
- Report the results of our searches for the use of these methods in transport and transport safety.
- Discuss bibliographic aspects of these methods in transport and transport safety.
- Assess how relevant are these methods for transport and transport safety.

Many researchers in transport and transport safety must from time to time have felt dissatisfaction over how methods in our field compare with those in medicine. The present paper will act as a resource documenting the extent to which high methodological standards have been adopted in our field to date, will assess how guilty we should feel about high methodological standards not being more common, and will suggest what we should advocate for the future.

What is meant by ‘evidence-based’?

In transport safety, things sometimes get changed (e.g., the speed limit might be reduced from 60 km/h to 50 km/h), and people want to know what the effect was. The same goes for transport more broadly, as when people want to know whether a campaign to encourage use of public transport rather than the car had any effect. Thus, what might be done is to measure the situation before the intervention, measure it afterwards, and make a comparison. To be careful, this before-after comparison might also be made at some other place where the change had not been introduced. What could possibly be wrong with such a procedure? Unfortunately, biases can easily creep in. The defence against biases that has become popular in medicine is the randomised controlled clinical trial. If allocation of experimental units to treatment or control groups really is random, then it has not been affected by anything capable of generating a bias. (And the likely size of random variation can be calculated mathematically.) But it is quite a leap from randomly assigning patients to aspirin or placebo groups to (say) the random conversion of junctions to roundabout control, or randomly making some bus routes free. Is that sort of thing at all practicable? The specific stimulus for this paper was an editorial by Fitz-Gibbon (2004) that is clear, to-the-point, and inspirational in its advocacy of higher methodological standards in social research.

Outline of randomised experimentation

Let us set out some features of how research should be conducted.

- Specify the dependent variable (the outcome) of interest.
- Identify the unit to which the intervention is being directed. For example, the units might be people (or in a transport example, intersections)
- Define all the units of interest, randomly choose some of them, and randomly assign each unit to either the treatment group or the control group.
- Measure the present condition of each unit.
- Apply the intervention to the units in the treatment group (and do nothing --- or nothing significant --- to those in the control group).
- Conceal, both from the people participating and from the researchers who are evaluating the outcomes, which units are in which of the groups. That is, conduct the research in a double-blind fashion.
- Measure the condition of each unit again (i.e., after the intervention has taken effect).
- The change of each unit is now known. The changes in the treatment group can now be compared with the changes in the control group.

Why is research done this way, or, at least, why do textbooks say research should be done this way? The background is that we typically wish to carry out a statistical test and say that an observed improvement was or was not statistically significant; or, at least, we wish to view the average change observed in the context of the natural variation that is present. Now, the research design sketched above incorporates a double comparison: there is a comparison of before the intervention with after, and there is a comparison of this (i.e., this difference, after minus before) between the treatment group and the control group.

- Suppose we did not have a control group. We might decide to treat all the experimental units, and compare before with after. The problem with this design is that it is quite common for things to happen that affect all the experimental units (e.g., there may be long term trends or specific events may occur); if so, we have no idea whether the observed change was due to the treatment we imposed or to the extraneous event. (And we may not even know about the extraneous event.)
- Suppose we did not have a longitudinal comparison. That is, we compare a group of units that already are in the treatment condition with a group in the untreated condition. In this case, the variation between units may be so great as to swamp any difference due to the factor of interest. Furthermore, one does not know why some units came to be in the treatment condition and others untreated --- for example, in health-related contexts, some general factor of concern about health may lead to some people both being more healthy and putting themselves into the treatment condition, and there is no causal connexion between treatment and health.

Research projects of a purely longitudinal or a purely cross-sectional type certainly are carried out, but the quality of evidence from them is widely regarded as greatly inferior to that from a control/treatment before/after comparison.

It should be mentioned that if there are some experimental units on which we have good reason to think that the experiment does not need to be performed (because we already are sure that the treatment is better, or the control condition is better, or that there is no difference), those units are excluded from the randomisation. The conclusions of the experiment then apply to the population from which the randomised units came (i.e., excluding the types of unit for which we already knew the answer). And if there are some experimental units that refuse to be randomised, they are excluded also. Provided this takes place before the randomisation, the experiment remains a valid one; however, the conclusions will apply only to the population of units that permit themselves to be randomised.

Systematic reviewing and meta-analysis

Randomisation (when doing the research) and systematic reviews (when weighing up research that has already been done) are the most distinctive features of the evidence-based approach.

The review process places a lot of emphasis on being thorough, including attempting to discover unpublished studies, and on weighting studies of high methodological quality much more heavily than those of lower quality. The basic idea in meta-analysis is to average results from several different investigations of the same topic --- if each is unconvincing because of a large standard error of the estimated effect, they may be convincing in total if they are all in the same direction. There are textbooks by Sutton, Abrams, Jones, Sheldon, and Song (2000) and Whitehead (2002).

Despite the thoroughness, and the objective combination of evidence from different studies, there is quite a long list of standard objections to the procedure --- that there may be publication biases (positive results may be more likely to be written up, and may be more likely to be accepted for publication when submitted), that the criteria that supposedly capture the quality of a study may not be valid, that the sampling frames (of the review, and of the studies included) may not quite correspond to the use we will make of the review, that the end result of the meta-analysis may depend on some statistical subtlety in the averaging process, that the principle of averaging results that may come from different populations (and may be statistically significantly different from one another) is a questionable one, and so on.

There are some features of transport and transport safety research that tend to make these objections rather more pertinent.

- ‘Grey’ literature (e.g., semi-published, perhaps not refereed, conference papers and research reports) is more important than in clinical medicine, and is likely to be more difficult to discover than journal articles.
- As we shall see below, the number of studies of high methodological quality on many issues is approximately zero. Thus, in weighing and combining the evidence, it becomes important to make distinctions at the low quality end of the scale. (a) Concerning before-after treatment-control comparisons, it is usually considered that such studies have some value if the control group really is an appropriate one. Their evidence may be almost valueless if that is not the case. No guidance is available to assist in making this judgment. (b) When there is no other evidence available, we might look at naive before-after comparisons. There is no general agreement about whether these are worth something or nothing.
- In medicine, a meta-analysis may be conducted on a topic on which dozens, perhaps hundreds, of studies have been conducted. In transport and transport safety, there are typically only a small number of relevant studies, and each of these has some good features and some poor.

- In transport and transport safety, the relevant studies may be more different from each other than they are in medicine.

Evidence-based social research

The drive for higher methodological standards has affected other fields, such as education, criminology, and social welfare. The advocates of randomised experimentation claim that it is by far the best way of determining what works and what doesn't work. In these and related fields of social science, the Campbell Collaboration (http://www.campbellcollaboration.org) is the equivalent of the Cochrane Collaboration. As to education, a recent document from the (United States) Coalition for Evidence-Based Policy (2003) strongly advocates using rigorous evidence to guide the choice and implementation of interventions in schools and classrooms. In a reflection of how far evidence-based arguments have come in this sector, this document targets not researchers in the field but the educational administrators who make the decisions on intervention implementation, and members of the public concerned about the decisions. The document is explicit in advising how to evaluate the evidence arising from research:

- Strong evidence requires both of the following.
  As regards quality, randomised controlled trials that are well designed and implemented.
  As regards quantity, the trials to be in at least two typical school settings, at least one of which must be similar to the proposed future application.
- Possible evidence would include randomised controlled trials that are good in quality and quantity but not reaching the standard required for strong evidence, and comparison group studies in which the intervention and comparison groups are very closely matched in relevant demographic and other variables.
- Otherwise, the intervention should be judged to be not supported by meaningful evidence. Included here will be all research using before-after comparisons, or using a comparison group that is not closely matched with the intervention group, plus meta-analyses that include such studies.

There is some guidance given as to the meaning of 'well designed and implemented'. This refers to matters such as using objective dependent variables, blinding of researchers to which participant has been allocated to which group, only a small proportion of participants missing when outcome data are collected, and so on. The document gives reasons why before-after studies and comparison group studies might fail to give the correct answer, and examples where they have failed in the past.

If we use this as a yardstick by which to measure the standard of transport research, we come up short. Most road safety and transport research does not reach the standard even of 'possible' evidence: typically, before-after comparisons are made with no control group, or with a control group that is only approximately suited to its purpose.

Most randomised experimentation in the criminological, educational, and social welfare fields seems to be from the United States. Presumably the reasons are the size of that country and the influence that Federal money can have. Riccio and Bloom (2001) give several examples of social experiments in the United States that involved randomisation. The units that were randomised were in most cases individuals. An exception was the Jobs-Plus Initiative, in which several public housing developments in each city were nominated, one of which was randomly selected to receive the intervention. The document from the Coalition for Evidence-Based Policy (2003) gives greatest attention to interventions for which individuals can be randomised, but it is plainly intended to apply to group randomisations also. The relevance of this to transport and transport safety is obvious --- while there may be some treatments that we could imagine randomising individual people to, for others the unit might be an intersection or a bus route or a whole suburb. For discussions on group randomisation, see Murray (1998) and Donner and Klar (2000).

As an example of evidence-based methods in a field (criminology) that must be as difficult as transport for randomised experimentation, Petrosino, Turpin-Petrosino, and Buehler (2003) conducted a systematic review of the 'Scared straight' programmes conducted in the United States. These involve juvenile delinquents visiting adult prisons and (perhaps) being so affected by that experience that they do not become inmates of such institutions in the future. They concluded that such programmes actually tended to increase delinquency. Incidentally, when the negative results from a programme of this type in California came out, the response was to end the evaluation, not the programme (Petrosino et al., 2003, p. 27), showing that society may not react to objective evidence in the way anticipated. Such programmes continue to be used, supported by evidence of an indirect nature (namely, what prisoners and the participants say about them). Petrosino et al. (2003) also note that shock value interventions are tried in other fields, including road safety, and wonder whether the results are as disappointing as with the 'scared straight' programmes. Oakley (2002) comments on this from the perspective of education research: 'These findings from research synthesis are worrying for fields such as education where so much of the 'evidence' is derived from small-scale qualitative research, depends heavily on practitioner judgements about the right thing to do, and/or is taken from poorly evaluated interventions.'

A pair of papers in the Journal of Economic Perspectives are complementary. That by Burtless (1995) broadly favours experimentation, whereas that by Heckman and Smith (1995) emphasises the problems and the alternatives. Useful edited collections are Mosteller and Boruch (2002) and Sherman (2003).

Transport and transport safety: the state of play

In the context of evidence-based everything, the terms 'high' and 'low' methodological quality are taken from controlled clinical trials: if a study does not involve randomised experimentation, then it is likely to be labelled as low quality. (See the report by the Coalition for Evidence-Based Policy, 2003, referred to above, and various research quality checklists.) We hope that is not considered offensive --- the discussion below will make clear that we think that transport and transport safety researchers have a good partial defence along the lines of 'It's the best that can be done in our field. Besides, randomised experimentation has its own set of problems'.

Earlier material

Well before the phrase 'evidence-based' began spreading through the world of research and evaluation, there were those in the transport sector who recognised the merits of stronger evidence.

- Helliar-Symons (1981), for example, evaluated yellow bar transverse carriageway markings at the approach to a roundabout. The bars decrease in separation as the roundabout is approached, creating a visual impression of speeding up, unless the driver reacts properly by slowing. Helliar-Symons examined whether this had had any effect on crash rates. Whilst not explicitly stating that assigning sites to test and control groups was wholly random, Helliar-Symons did avoid using crash record as a selection criterion, because of the problem of regression towards the mean.

- Nichols, Ellingstad and Reis (1981) documented the evaluations of drink driving education and treatment programs through the 1970's in the United States, producing a review of the various studies and their methods and noting the progression towards the end of the decade to randomisation.
- McKnight and Edwards (1982) conducted an evaluation of written driving manuals and tests in Virginia, United States, with random assignment and control groups to examine the effect on crash rates. Significantly, their final conclusion is a call for further research in this area to use large sample sizes.
- Fosser (1992) evaluated the effect of periodic vehicle inspection on crash rates in Norway, with a randomised controlled experiment that was devised in the light of the methodological weaknesses of previous studies in the field.

It may be difficult, in transport, even to achieve the elementary requirement that if you want to know the effect of intervention I, then that intervention alone must be made. If there is a package of interventions I, J, and K that are all made at once, it is impossible to disentangle their separate effects. The Stevenage Superbus experiment (Buckles, 1975) was a case where a serious attempt was made to implement different interventions (to increase patronage, via service improvements and fare reductions) at different times, though it was not possible to follow that strategy fully.

These are examples and are no doubt not the only early efforts to evaluate and review with evidence based principles. But they suffer from a lack of 'branding' in their titles and abstracts of the words and phrases that indicate evidence based literature and so their citations in bibliographic databases are highly unlikely to have been indexed with such terms. A fuller discussion of information retrieval concerns follows later in this paper.

Cochrane reviews

The Cochrane Collaboration (http://www.cochrane.org) commissions and makes available in electronic and hard copy formats systematic reviews of interventions in health-related areas, not only medicine. Several Cochrane reviews have been devoted to transport safety issues (see Table 1). The majority have covered driver and other road user education interventions but there have also been reviews of traffic calming, bicycle helmets and pedestrian and cyclist visibility measures. A review of injury prevention for problem drinkers includes driving-related injuries. There are also protocols for reviews currently underway (also listed in Table 1) which cover similar topics but with several more studies in the traffic arena, covering issues such as speed enforcement and red light cameras.

The completed reviews were conducted and reported in the standard Cochrane format of documented search strategy (electronic databases, manual checking of reference lists, consultation with experts in the field), identification of randomised controlled trials (and in some cases controlled trials), meta-analysis of methods, comparisons, and conclusions. These reviews show that evidence-based transport safety research is in its infancy. From the trials that were initially located, very few were identified that had been conducted to good methodological standards (see Table 2). For example, in their review on safety education for pedestrians, Duperrex, Roberts and Bunn (2004) could only locate 15 randomised controlled trials conducted between 1976 and 1997, and the general quality of these studies was rated as poor.

There is overlap of authorship between nearly all of the studies. Evidently there is a small core group promoting the evidence-based methods, and not many other people.

Table 1 Titles of Cochrane reviews of transport safety issues, and those in preparation (from http://www.cochrane.org/cochrane/revabstr/mainindex.htm)

Area-wide traffic calming for preventing traffic related injuries. F Bunn et al. (See also Bunn et al. (2003))
Helmets for preventing head and facial injuries in bicyclists. D C Thompson et al. (There are comments disputing the conclusions of this review at http://www.update-software.com/comcritusers/)
Interventions for increasing pedestrian and cyclist visibility for the prevention of death and injuries. I Kwan and J Mapstone. (See also Kwan and Mapstone, 2004.)
Interventions for preventing injuries in problem drinkers. T Dinh-Zarr et al.
Post-licence driver education for the prevention of road traffic crashes. K Ker et al.
Safety education of pedestrians for injury prevention. O Duperrex et al.
School based driver education for the prevention of traffic crashes. I Roberts et al. (See also Cochrane Injuries Group Driver Education Reviewers, 2001.)

Cochrane Protocols (systematic reviews currently being undertaken):

Alcohol ignition interlock programs for reducing drink driving recidivism
Graduated driver licensing for reducing motor vehicle crashes among young drivers
Helmets for preventing injury in motorcycle riders
Increasing motorcycle and rider conspicuity for preventing death and injury in motorcyclists
Interventions for promoting use of booster seats for children aged 4-8 traveling in cars
Interventions in the clinical setting for preventing unintentional injuries among children and teenagers aged 0-19 years
Non-legislative interventions for the promotion of cycle helmet wearing by children
Red light cameras for the prevention of road traffic crashes
Speed enforcement detection devices for preventing road traffic injuries
The ‘WHO safe communities’ model for the prevention of injury in whole populations

Table 2

Numbers of trials identified as satisfying the inclusion criteria for several systematic reviews relevant to transport safety; for details of these reviews, see http://www.cochrane.org

Notes:
(1) Grounds on which reviewers are likely to class trials as poor in methodological quality include: inadequate concealment of allocation, failure to blind when assessing outcome, loss of many subjects to follow-up, failure to report outcome measures of the relevant type, failure to implement the experimental condition as intended.
(2) These seven systematic reviews were not by seven unrelated groups of researchers. Except for the group responsible for the Thompson et al. review, the others are connected by common members.

Bunn et al. aimed to evaluate the traffic safety benefits of area-wide traffic calming. They searched for randomised controlled trials and controlled before-after studies.

No randomised controlled trials were identified. 16 controlled before-after trials were identified (14 in western Europe, 2 in Australia).

Thompson et al. aimed to determine whether helmets for bicyclists tend to reduce head, brain, and facial injuries. They searched for controlled studies evaluating helmet use in bicyclists who had a crash.

No randomised controlled trials were found. 5 well-conducted case-control studies were found.

Kwan and Mapstone aimed to quantify the effect of aids to pedestrian and cyclist conspicuity on accidents and on drivers' responses. They searched for randomised controlled trials and controlled before-after trials.

Accidents: no trials were found.

Drivers' responses (such as reaction times) in experimental settings: 37 trials were found.

Dinh-Zarr et al. aimed to assess the effect of interventions for problem drinking (i.e., alcohol dependence, alcohol abuse, or hazardous use of alcohol) on subsequent injury risk. They searched for randomised controlled trials.

4 trials were found that assessed the effect on motor vehicle trauma.

Ker et al. aimed to quantify the effectiveness of post-licence driver education in reducing crashes. They searched for randomised controlled trials.

24 trials were found (23 were in the U.S.A.), 20 being remedial in nature and 4 being advanced. However, Ker et al. considered their methodological quality to be poor.

Duperrex et al. aimed to quantify the effectiveness of pedestrian safety education programmes. They searched for randomised controlled trials.

Accidents: no trials were found that used accidents as the outcome measure.

15 randomised controlled trials were found that used some other outcome measure, such as observed behaviour. However, Duperrex et al. considered their methodological quality to be poor (poor concealment of allocation, inadequate blinding of outcome assessment).

Roberts et al. aimed to quantify the effect of school-based driver education on licensing and accidents. They searched for randomised controlled trials.

3 trials were found that used accidents as the outcome. (These were conducted between 1982 and 1984, in the U.S.A., Australia, and New Zealand.) 2 of these also examined the effect on licensing.

Table 3. Some traffic engineering papers that show the influence of evidence-based medicine

Elvik (1995) --- lighting
Elvik (1995) --- guardrails and crash cushions
Burke et al. (1996) --- modification to bus stop
Elvik (1997) --- blackspot treatment
Vaa (1997) --- speed limit reduction
Elvik and Mysen (1999) --- accident reporting
Elvik (2001) --- traffic calming
Elvik et al. (2001) --- bypasses
Towner et al. (2001) --- children (including traffic calming)
Retting et al. (2002) --- traffic signal retiming
Egan et al. (2003) --- road building and health
Elvik (2003) --- roundabouts
Elvik and Greibe (2003) --- porous asphalt and safety
Flannery and Maccubbin (2003) --- red light cameras
Retting et al. (2003) --- traffic engineering measures for pedestrians
Ulleberg (2003) --- motorcyclists (including road environment)
van Driel et al. (2004) --- edgelines

Systematic reviewing sometimes leads to conclusions that contradict conventional opinion. Because of the emphasis on studies of high methodological rigour, and the objective way in which systematic reviewing is conducted, the people who do this may feel they have discovered a truth that has not been revealed to others, and combat may ensue. For example, in our own field, the Cochrane Injuries Group Driver Education Reviewers (2001) reviewed the effect of school driver education on licensing and crashes, restricting attention to randomised controlled trials, and concluded that the (three) studies found provide no evidence that driver education reduces road crash involvement; indeed, it may worsen things, by leading to earlier licensing. Such a finding is contrary to the (U.K.) government policy promoting driver education in schools. For a more straightforward medical example, we may mention the controversy over whether mammographic screening for breast cancer is beneficial. Who is qualified to weigh up the evidence, a cancer expert or an expert in meta-analysis? It seems to be the case that despite the supposedly objective methods of systematic reviewing, it is not controversy-free.

Other papers influenced by the evidence-based approach

Table 3 lists some traffic engineering papers that used methods similar to those accepted by the Cochrane reviews, though our impression is that most of these were not as strict about what constitutes good methodology as the Cochrane Collaboration is. (Experimentation that is not intervention in the real world, such as that in psychological, physiological, engineering, and other laboratories, is outside our scope, and so are trials of different medical treatments of the consequences of accidents.) There have been many meta-analyses published in the public health literature, some of which were able to include numerous studies that involved the randomisation of individual people. Especially noteworthy are:

- Supplement 1, Volume 16 (1999), of the American Journal of Preventive Medicine, which included introductions and comments by Rivara, Thompson, Beahler, and MacKenzie (1999), Williams (1999), and Wagenaar (1999), and papers on: Effectiveness of health promotion programs to increase motor vehicle occupant restraint use among young children; Evaluating interventions that promote the use of rear seats for children; Effectiveness of primary and secondary enforced seat belt laws; Effects of high school driver education on motor vehicle crashes, violations, and licensure; Effectiveness of graduated driver licensing in reducing motor vehicle crashes; The effect of random alcohol screening in reducing motor vehicle crash injuries; The specific deterrence of administrative per se laws in reducing drunk driving recidivism; Evaluation of the effectiveness of low blood alcohol concentration laws for younger drivers; Effectiveness of ignition interlock devices in reducing drunk driving recidivism.
- Supplement 4, Volume 21 (2001), of the American Journal of Preventive Medicine, which included Zaza et al. (2001), and papers on: Reviews of evidence regarding interventions to increase use of child safety seats; Reviews of evidence regarding interventions to increase the use of safety belts; Reviews of evidence regarding interventions to reduce alcohol-impaired driving.
- A review of 28 systematic reviews in traffic safety by Morrison, Petticrew and Thomson (2003).
- Towner, Dowswell, MacKereth and Jarvis (2001) examine the quality of evidence of a range child injury prevention measures, noting randomised trials in the areas of road safety education and interventions to increase use of helmets and restraints.

There have been meta-analyses in the economics literature, too --- for example, on the valuation of traffic noise, air quality, and human life, and elasticities of public transport demand and fuel demand.

Rivara et al. (1999, p. 4) noted that ‘randomized controlled trials have become the standard study type by which the effectiveness of therapy for the treatment of medical problems is evaluated’, but lamented that ‘it is very difficult to conduct randomized controlled studies in which the intervention is a law’. For the motor vehicle interventions of interest to them, few randomised trials had been conducted, though there had been many controlled trials (the groups being assigned by the investigator, albeit not randomised). Commenting in the same special issue of the American Journal of Preventive Medicine, Williams (1999, p. 1) states, ‘What stands out in these reviews is the paucity of intervention evaluations.... A related theme concerns the lack of high-quality evaluations.’ And Wagenaar (1999, p. 9) concurs, ‘I find the most interesting implications from the reviews reported here related to research methods used and how the results are reported. Note the large number of documents screened to find a few with useful information.... 54708 papers and reports were identified.... only 161 survived initial screening and were included in the ten review papers.... The pessimist might conclude that there are just too many useless papers published, and too many incompetently conducted studies.’

Table 4. Methodological aspects of studies of the effects of new roads considered in Egan et al. (2003)

(a) Effects on injuries

Control for general trends
Reliable / representative sample of data
Sufficient data presented to validate results
Control for regression to the mean
Assessment at least 3 years before and 3 years after
Compares more than one new road
Injury severity considered
Number of individual casualties included
Accident migration across wider road network considered

(b) Disturbance of residents

Appropriate sampling
Response rate / follow-up > 60%
Controls / adjustment for confounders
Appropriate exposure measures
Adaptation to disturbance considered
Impact on secondary roads considered
Sufficient data presented to validate results
Compared more than one new road
Prospective study

Egan, Petticrew, Ogilvie and Hamilton (2003) conducted a systematic review of the health impacts of new roads. They summarised studies on such matters as the effects of new roads on injuries and of major urban roads and new bypasses on disturbance to local residents and among residents of the bypassed area. They noted a number of methodological aspects of the studies they reviewed. These may be of general interest, perhaps almost a checklist of good methodology, and we have listed them in Table 4.

Forbes (2003) has an appendix, 'Evidence-based road safety'. By 'evidence-based', he means the same thing we do, noting this 'is not the tacit acceptance of conclusions from poorly conducted research, simply because it is published, reported in a trade journal or presented at a technical conference' (p. 185). The report deprecates the low methodological standard of much road safety research: 'The knowledge base with respect to the safety impacts of traffic operations and central strategies is underdeveloped... Until sufficient good quality research has been conducted and made available... the practitioner will likely have to rely on prejudice, hunch and guesswork' (p. 190). Much of the document is taken up with reviewing the safety effects of traffic engineering measures --- and Forbes found almost all of the research projects were methodologically unsatisfactory. Thus if Forbes had followed his definition of evidence-based research, almost all the projects he reviews should have been omitted.

The meta-analysis by Wells-Parker, Bangert-Drowns, McMillen and Williams (1995) was considered a model example by several commentators. But it is plain why other meta-analyses do not reach the same standards. Wells-Parker et al. were dealing with remediation of drinking/driving offenders. These are individual people, who can be randomised to one group or another in a way familiar to medical researchers. Thus numerous studies of that type had been carried out, and plenty remained after restricting consideration to those of high quality. For many other topics in transport and transport safety, the unit of intervention and randomisation is not a person, and thus very few (if any) randomised experiments have been conducted. In the meta-analysis, the choice is then between excluding studies of low methodological quality, and being left with nothing, or basing the averaged results on the low quality studies, all of which might possibly share the same biases.

In their recently published book, Elvik and Vaa (2004) offer several relevant observations on the methodological strengths of road safety measure evaluations.

- From their chapter on road design and road furniture (pp. 260-262): 'The majority of studies of the effects of road layout and road equipment on accidents tend not to be based on a random sample of sites drawn from a known population or sampling frame. This is a basic weakness of studies in this area. Strictly speaking, this means that the results of many studies cannot be generalised to places and conditions other than precisely those for which they were carried out. The population to which results are meant to apply is, in many cases, not clearly defined.... In order to claim that a particular measure is a cause of changes in the accident figures, one must rule out that these changes are due to other events or factors. Strictly speaking, such a requirement can only be fulfilled in experiments. In non-experimental studies, it can never be totally excluded that changes in the accident figures were caused by uncontrolled confounding factors. Only one of the measures has been studied experimentally. This is the use of game mirrors designed to prevent accidents involving wild animals and game. For all other measures affecting road layout and road equipment, the results presented come from more or less well-controlled non-experimental studies.'
- They note that experiments or good quasi-experimental studies have been conducted on (a) salting of roads (p. 408), (b) speed limits, road markings, area-wide traffic calming, and signal-controlled pedestrian crossings (p. 465), (c) anti-lock brakes, extra-high mounted stop lamps, daytime running lights on cars, reflective materials and protective clothing, seat belts in cars, vehicle crashworthiness, and regulating motor power of mopeds and motorcycles (p. 614), (d) vehicle inspection (pp. 802-803), (e) basic driver training and the treatment of problem drivers (p. 832), (f) information and campaigns (p. 940), and (g) police enforcement and sanctions (p. 964).

Comments

The general difficulty of making a search for methodology across a wide subject area, along with the prevalence of grey literature in transport and transport safety, means that we are sure to have missed some relevant studies. Even so, it seems safe to conclude that randomised experimentation is rare in transport and transport safety. Such studies as there have been tend to be concentrated in public health. Randomisation of units other than people is rare. (Regarding this, see also the Section on evidence-based social research, above.) Blinding to treatment or outcome of either participants or researchers is generally so impracticable it is hardly discussed.

Meta-analysis, on the other hand, has been much more enthusiastically embraced. However, our impression is that many meta-analyses have included studies of quite low methodological quality, because there are so few randomised experiments in our field. They are thus open to the 'garbage in, garbage out' objection. We hasten to repeat that we do not hold the view that all research that does not use randomised experimentation is garbage.

For us, the key points that have come out of the literature have been:

- There are some people dissatisfied with transport and transport safety research for not making greater use of randomised experimentation, and openly criticising the field.
- On some topics, especially those close to medicine, randomisation of individual people is possible and is common.
- Randomisation and treatment of units other than individual people is rare.
- Randomisation and treatment of other units does sometimes happen. For example, Burke, Lapidus, Zavoski, Wallace and Banco (1996) randomised bus stops, and Retting, Chapline, and Williams (2002) randomised intersections. And sometimes groups of people (e.g., a school) are randomised when the treatment is one of individuals.
- Meta-analysis is common, but often is not fully followed through, in the sense that studies of quite low methodological quality are retained in the sample.

Bibliographical limitations of the present study

Our search of bibliographic databases utilised ATRI, TRANSPORT, ERIC, PsychINFO, MEDLINE, ENGINE and Ei Compendex Web. These are citation databases which index papers in journals and conference proceedings and research reports in the fields of transport, education, psychology, medicine, and engineering. Search terms included the topics of randomisation, controlled trials, meta-analysis and evidence-based research. In addition, a search of the World Wide Web was undertaken, making sure that resources that are not spidered by search engines were investigated independently. (These include large and specialist library catalogues and shared catalogues such as TRANSCAT as well as the Cochrane Library database.) Reference lists from located resources were also manually checked. In order to be able to have full information on matters such as research methodology, the search concentrated on English language material.

A state-of-play analysis such as this requires a certain level of qualification as to its completeness, for there are some hurdles that are encountered in locating material. Most of these issues apply to some degree to any systematic reviewer in the transport and transport safety fields, and highlight the differences that exist between medicine and transport. They have however been more significant for this paper, as it attempts to document evidence-based activities across the transport sector, rather than focussing on a single intervention or road user type. Even the phrase 'evidence-based' has been an issue, with many papers not actually coming within the scope of what we mean. Marston and Watts (2003) give some attention to the dilution of the meaning of the phrase and Dopson, Locock, Gabbay, Ferlie and Fitzgerald (2003) highlight some confusions that occurred within the medical profession over what constituted evidence-based medicine. Scanning titles and abstracts in conducting our search for material showed that awareness of the issue and clarity of definition are not as developed as in medicine. And of course we should note that papers and reports that used similar methods but pre-dated the current phrases ('evidence-based', 'randomised control', etc.) are difficult to locate even with the electronic tools available to researchers and information specialists today.

Systematic reviewing in transport and transport safety

Grey literature has always been important in transport and transport safety, and the rise of online publishing has led to a rapid growth in reports and papers being produced and disseminated by small research centres and individual researchers. Such sources were not traditionally included in databases, though more and more of these are now being included. In medicine, evidence-based research results are much more likely to be published in journals which are unlikely to miss inclusion. Difficulties also extend to material that is in databases. Wentz, Roberts, Bunn, Edwards, Kwan and Lefebvre (2001) highlight the difficulties in conducting road safety systematic reviews given the current standards of indexing in the field. They urge authors to take special care to document the study design in their article or report, and for information specialists to carefully and consistently index study methodology in road safety databases. Egan et al. (2003) found a much greater proportion of suitable studies by consulting experts in the field and manually checking reference lists than from electronic sources, and this is not an isolated example.

Would-be systematic reviewers are likely to find that for many transport safety topics the evidence base is too small to make meaningful judgments on intervention efficacy. This is not really a criticism of the sector. There have been researchers looking at evidence-based principles for some time, as the existence of earlier published material shows. And, while evidence-based medicine has been widespread for some time, reviews in that area still show significant gaps. For example, Dickinson, Bunn, Wentz, Edwards and Roberts (2000) found a distinct lack of sizeable and methodologically sound trials in the area of head injury.

Disadvantages of randomised experimentation

There is an influential body of opinion that says that the only valid way of finding out what works is randomised experimentation. We have confirmed what others have said, that randomised experimentation is very rare in transport and transport safety. Can we therefore conclude that researchers have therefore been wasting their time for the past 50 years? In our view, no. Those sentences do not take account of the real practical difficulties with randomised experimentation, the principled objections to randomised experimentation, and the merits of conventional forms of research. Of these, the objections to experimentation are the most interesting issue. Even if we know that randomised experiments are difficult to do, and even if we feel that what we do instead has real worth, we may feel guilty about not doing the best form of research --- unless we know that randomised experimentation has its own problems.

A number of principled criticisms of randomised controlled trials will be made below. Many of them have been considered in the medical and social science literatures. By way of overview, our impression is that all are intellectually respectable, and there is not a knock-out response to any of them. That is, they all deserve to be taken seriously, and may or may not constitute good arguments, depending upon the particular issue being considered. If that is right, we cannot simply say that randomised experimentation is the one and only correct way of doing research.

In medicine, the stimulus (e.g., half an aspirin) may be obviously the same everywhere. Social treatments (including those in transport and transport safety) are not so standardised. Let's suppose that some intervention (e.g., 'traffic calming') can be done in several different ways. Let's suppose a randomised trial shows it to be ineffective. Let's suppose our local traffic engineer says, 'But I know how to do it properly, and I really will get safety gains'. Is the randomised trial really evidence against our engineer? Our view is that it may not be. It is good evidence against the variety of traffic calming that was used in the trial. Extension of the conclusion to a different variety is a matter of judgment, and it is not obvious that this will outweigh the arguments of our local traffic engineer. Another version of this point would be that there is only one intersection (or town, or bus route, or model of car) that has its own specific characteristics, that no previous randomised trial is relevant, and that the proposed intervention is designed for it. To support this view, we note that the Cochrane Injuries Group Driver Education Reviewers (2001) drew certain conclusions from trials of driver education programmes in Australia, the U.S.A., and New Zealand between 1982 and 1984. But they did not immediately say those conclusions were valid for a particular driver education programme in the U.K. in 2001. Rather, they weighed up the characteristics of different programmes, and made a judgment concerning the relevance of the conclusions.

Thus the difficulty of standardising the intervention may be an issue. So may be the appropriateness (or otherwise). It may be felt that it ought to be possible to adapt an intervention to the particular circumstances of a particular place. If it really is standardised, there is no opportunity to do that.

Indeed, the specification of the intervention that we are interested in needs to include the degree to which it is standardised. One that is highly standardised is different from one that involves a considerable degree of local latitude. For example, a 'traffic calming' intervention might involve the local planners and engineers going on a course to understand it, being provided with textbooks and checklists and objective criteria, and being supervised by an accredited expert. Or it might involve them being given merely a general strategy and a few suggestions about speed humps and roundabouts. These are different interventions, and conclusions from one would not necessarily apply to the other.

Even if the intervention can be closely specified so that it is the same everywhere, it may be impossible to specify its context or environment. The effect of the intervention may depend on social, economic, and geographical circumstances. In medicine, this argument turns up in the form that the results of a randomised trial may apply on the average, but it is difficult to judge if they apply to a specific patient (Feinstein and Horwitz, 1997). That is, the inclusion criteria for the trial may have been broad in respect of a variable that the clinician suspects may be important, and which is known for this patient; the clinician would prefer to have results for a group that are similar to the patient in this respect. This in no way contradicts an argument that the trial is not relevant to today's patient because the inclusion criteria in respect of some other variable may have been sufficiently narrow to exclude him or her.

Putting this in other words, there may be an interaction of the intervention with the place where it is made. If we in Adelaide decide to accept the conclusions of a high-quality research project from Los Angeles (say), we are saying that the difference in conditions will not affect the effect of the intervention. (We might not know what conditions might be relevant. But an example is the dependent variable of interest, whatever that is. If this is accident rate, for example, to transfer a conclusion from one place to another that has a different accident rate involves assuming that accident rate does not have a big impact on the conclusion.) Certainly we are not compelled to accept the conclusions from Los Angeles: among the inclusion criteria for the experimental units was that they were in Los Angeles, and ours in Adelaide do not satisfy that. And if we in Adelaide prefer to accept the conclusions of a high-quality research project from Los Angeles rather than those of a less-rigorous research project from Perth (say), we are saying difference in conditions from Los Angeles will have a lesser effect than the methodological deficiencies. These seem to be matters of judgment, not ones that can be resolved by universally-accepted rules.

The process of adopting explicit criteria for the units to be included in the experiment and then randomising them to one group or another ensures that the conclusions have internal validity, but there may be a question mark over their external validity and generalisability. There is typically a careful definition of the units (e.g., patients) who are permitted to participate in the trial, with an explicit list of inclusion and exclusion criteria. Conclusions will only apply only to this population. Any attempt to extend the conclusions to units having other characteristics is a matter of judgment, not something that can be said to be evidence-based. Further, some units may refuse to enter the trial. There may be concern that refusal is not a random matter, but is affected by some characteristics of the units. Conclusions will apply only to the type of unit that was offered participation in the trial and accepted.

Discussion

We fear that many issues in transport safety and transport will raise insurmountable problems for randomisation. Nevertheless, it does seem to us that in the case of some issues, randomisation ought to get more consideration than it does at present. But, because randomised experimentation is so difficult, it is likely that the process of listing the alternative methodologies, considering their merits, conducting pilot studies, identifying the problems, and judging whether the problems can be overcome --- in short, of planning the research --- will need to take longer and consume more resources than it typically does at present. Thus researchers and those funding research need to consider how this will be paid for. And this is not only a matter of finding the dollars: there may be a difference in the mentality and attitudes of a researcher and a research planner!

Admitting the difficulties, randomised experimentation does have unique advantages. These need greater publicity than they receive at present --- more opportunities for rigorous research, and stronger attempts to overcome the difficulties, might emerge if there were wider appreciation of the benefits. There is a case, too, for closer cooperation between several jurisdictions on a given research question, and collaboration in large-scale experiments. Admittedly this will involve surrender of autonomy, and not being able to adapt the intervention to local conditions and circumstances.

Research projects need to be large enough to lead to a reasonably small standard error of the estimated effect size. Perhaps more money needs to be spent on experimental interventions themselves, to make sure they are big enough to have a detectable effect. As to small interventions, an estimate of their effect is likely to have a large standard error associated with it. It is to be hoped that they are documented adequately, so that one day all those of similar type can be included in a meta-analysis.

The starting point for this paper was:

- The assault by the enthusiasts for evidence-based everything on old ways of doing research, and their claims that their methods generate evidence that is much more credible than any other.
- Common sense objections to those claims, along the lines of 'Your high-quality evidence about X is not relevant to me, because the way I will implement X is different, or my city is different', and additional problems with practicalities of randomisation, as the relevant units are often not individual people.
- It is worthwhile carefully examining evidence-based everything, to determine if there is some flaw that limits its relevance in road safety and transport. And, on the other side of the argument, the objections may be weak excuses for keeping things the way they always have been; is there really any substance to them?

Our present view is:

- The methods of evidence-based everything --- notably, randomised experimentation and systematic literature review --- do generate uniquely valid and convincing evidence. This is not some sort of con-trick by snake-oil salesmen.
- The principled objections to evidence-based everything cannot be dismissed out of hand. The enthusiasts for randomised experimentation have not answered them comprehensively, they do have substance. The practical difficulties are real, too.
- Consequently, randomised experimentation should be considered more seriously than it usually has been in the past. It is likely that in most instances it will be rejected, either because of principled objections or impracticability, and some methodology of 'lower quality' chosen instead, but that will not be known beforehand.
- There will be consequences for the organisation of research and the relationships between those paying for it and those doing it.

There are many questions that randomisation and meta-analysis do not give the answers to --- what outcome variable to choose, what summary statistic to calculate, what statistical test to perform, and so on. Enthusiasts for the evidence-based approach will say of course it will not answer these, as they are not the questions it is addressing. That reply is fair, but raising such questions does dent the credibility of the more extreme claims about the usefulness of the evidence-based approach.

Acknowledgements

The Centre for Automotive Safety Research receives core funding from both DTUP and the South Australian Motor Accident Commission.

The views expressed in this report are those of the authors and do not necessarily represent those of the University of Adelaide or the sponsoring organisations.

References

Buckles P (1975) *The Stevenage Superbus experiment* Stevenage: Stevenage Development Corporation

Bunn F, Collier T, Frost C, Ker K, Roberts I and Wentz R (2003) Traffic calming for the prevention of road traffic injuries: systematic review and meta-analysis *Injury Prevention* 9(3), 200-204

Burke GS, Lapidus GD, Zavoski RW, Wallace L and Banco LI (1996) Evaluation of the effectiveness of a pavement stencil in promoting safe behavior among elementary schoolchildren boarding school buses *Pediatrics* 97(4), 520-523

Burtless G (1995) The case for randomized field trials in economic and policy research *Journal of Economic Perspectives* 9(2), 63-84

Coalition for Evidence-Based Policy (2003) *Identifying and implementing educational practices supported by rigorous evidence: a user friendly guide* Washington DC: Institute of Education Sciences, US Department of Education

Cochrane Injuries Group Driver Education Reviewers (2001) Evidence based road safety: the driving standards agency's schools programme *The Lancet* 358(9277), 230-232. (This is reproduced in *Proceedings of the 3rd international interdisciplinary evidence-based policies and indicator systems conference* (2001) 8-13. Durham: Curriculum, Evaluation and Management Centre, University of Durham)

Dickinson K, Bunn F, Wentz R, Edwards P and Roberts I (2000) Size and quality of randomised controlled trials in head injury: review of published studies *British Medical Journal* 320(7245), 1308-1311

Donner A and Klar N (2000) *Design and analysis of cluster randomization trials in health research* London: Arnold

Dopson S, Locock L, Gabbay J, Ferlie E and Fitzgerald L (2003) Evidence-based medicine and the implementation gap *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine* 7(3), 311-330

Duperrex O, Roberts I and Bunn F (2004) Safety education of pedestrians for injury prevention (Cochrane Review). *The Cochrane Library* Issue 1, 2004 Chichester: Wiley

Egan M, Petticrew M, Ogilvie D and Hamilton V (2003) New roads and human health: a systematic review *American Journal of Public Health* 93(9), 1463-1471

Elvik R (1995) Meta-analysis of evaluations of public lighting as accident countermeasure *Transportation Research Record* No 1485, 112-123

Elvik R (1995) The safety value of guardrails and crash cushions: A meta-analysis of evidence from evaluation studies *Accident Analysis and Prevention* 27(4), 523-549

Elvik R (1997) Evaluations of road accident blackspot treatment: A case of the Iron Law of evaluation studies? *Accident Analysis and Prevention* 29(2), 191-199

Elvik R (2001) Area-wide urban traffic calming schemes: A meta-analysis of safety effects *Accident Analysis and Prevention* 33(3), 327-336

Elvik R (2003) Effects on road safety of converting intersections to roundabouts: Review of evidence from non-US studies. *Transportation Research Record* No 1847, 1-10

Elvik R, Amundsen FH and Hofset F (2001) Road safety effects of bypasses *Transportation Research Record* No 1758, 13-20

Elvik R and Greibe P (2003) *Safety aspects related to low noise road surfaces* Report 680/2003 Oslo: Institute of Transport Economics

Elvik R and Mysen AB (1999) Incomplete accident reporting: Meta-analysis of studies made in 13 countries *Transportation Research Record* No 1665, 133-140

Elvik R and Vaa T (eds) (2004) The handbook of road safety measures Amsterdam: Elsevier

Feinstein AR and Horwitz RI (1997) Problems of the 'evidence' of 'evidence-based medicine' *American Journal of Medicine* 103(6), 529-535

Fitz-Gibbon C (2004) Editorial: the need for randomized trials in social research *Journal of the Royal Statistical Society, Series A* 167(1), 1-4

Flannery A and Maccubbin R (2003) *Using meta-analysis techniques to assess the safety effects of red-light running cameras* Paper presented at the 82nd Annual Meeting of the Transportation Research Board http://www.ltrc.lsu.edu/TRB_82/TRB2003-000512.pdf (online accessed 7/06/04)

Forbes G (2003) *Synthesis of safety for traffic operations* Ottawa: Transport Canada http://www.intus.ca/synthesis/Synthesis_Final_Report.pdf (online accessed 28/05/04)

Fosser S (1992) An experimental evaluation of the effects of periodic motor vehicle inspection on accident rates *Accident Analysis and Prevention* 24(6), 599-612

Heckman JJ and Smith JA (1995) Assessing the case for social experiments *Journal of Economic Perspectives* 9(2), 85-110

Helliar-Symons RD (1981) *Yellow bar experimental carriageway markings --- accident study* Laboratory Report 1010 Crowthorne: Transport and Road Research Laboratory

Kwan I and Mapstone J (2004) Visibility aids for pedestrians and cyclists: a systematic review of randomised controlled trials *Accident Analysis and Prevention* 36(3), 305-312

Marston G and Watts R (2003) Tampering with the evidence: a critical appraisal of evidence-based policy-making *The Drawing Board: An Australian Review of Public Affairs* 3(3), 143-163

McKnight AJ and Edwards R (1982) An experimental evaluation of driver license manuals and written tests *Accident Analysis and Prevention* 14(3), 187-192

Morrison DS, Petticrew M and Thomson H (2003) What are the most effective ways of improving population health through transport interventions? Evidence from systematic reviews *Journal of Epidemiology and Community Health* 57(5), 327-333

Mosteller F and Boruch R (eds) (2002) *Evidence matters: randomized trials in education research* Washington, DC: Brookings Institution Press.

Murray DM (1998) *Design and analysis of group-randomized trials* New York: Oxford University Press

Nichols JL, Ellingstad VS and Reis RE (1981) The effectiveness of education and treatment programs for drinking drivers: a decade of evaluation pp 1298-1328 of Goldberg L (ed) *Proceedings, 8th International Conference on Alcohol, Drugs and Traffic Safety* 3 Stockholm: Almqvist & Wiskell

Oakley A (2002) Social science and evidence-based everything: the case of education *Educational Review* 54(3), 277-286

Petrosino A, Turpin-Petrosino C and Buehler J (2003) 'Scared straight' and other juvenile awareness programs for preventing juvenile delinquency. Campbell review update I *The Campbell Collaboration Reviews of Intervention and Policy Evaluations (C2-RIPE)* Philadelphia: Campbell Collaboration

Retting RA, Chapline JF and Williams AF (2002) Changes in crash risk following re-timing of traffic signal change intervals *Accident Analysis and Prevention* 34(2), 215-220

Retting RA, Ferguson SA and McCartt AT (2003) A review of evidence-based traffic engineering measures designed to reduce pedestrian - motor vehicle crashes *American Journal of Public Health* 93(9), 1456-1463

Riccio JA and Bloom HS (2001) *Extending the reach of randomized social experiments: new directions in evaluations of American welfare-to-work and employment initiatives* Working Paper on Research Methodology New York: Manpower Demonstration Research Corporation

Rivara FP, Thompson DC, Beahler C and MacKenzie EJ (1999) Systematic reviews of strategies to prevent motor vehicle injuries *American Journal of Preventive Medicine*, 16(S1), 1-5

Sherman LW (ed) (2003) Misleading evidence and evidence-led policy: Making social science more experimental *Annals of the American Academy of Political and Social Science*, vol 589

Sutton AJ, Abrams KR, Jones DR, Sheldon TA and Song F (2000) *Methods for meta-analysis in medical research* Chichester: Wiley

Towner E, Dowswell T, MacKereth C and Jarvis S (2001) *What works in preventing unintentional injuries in children and young adolescents? An updated systematic review* London: Health Development Agency

Ulleberg P (2003) *Motorcycle safety --- A literature review and meta-analysis* (In Swedish) Report 681/2003 Oslo: Institute of Transport Economics

Vaa T (1997) *Speed limit reduction in built-up areas: Effects on speed and accidents* (In Norwegian) Report 1085/1997 Oslo: Institute of Transport Economics

van Driel CJG, Davidse RJ and van Maarseveen MFAM (2004) The effects of an edgeline on speed and lateral position: A meta-analysis *Accident Analysis and Prevention* 36(4), 671-682

Wagenaar AC (1999) Importance of systematic reviews and meta-analyses for research and practice *American Journal of Preventive Medicine* 16(S1), 9-11

Wells-Parker E, Bangert-Drowns R, McMillen R and Williams M (1995) Final results from a meta-analysis of remedial interventions with drink/drive offenders *Addiction* 90(12), 907-926 (with commentaries, 1587-1601)

Wentz R, Roberts I, Bunn F, Edwards P, Kwan I and Lefebvre C (2001) Identifying controlled evaluation studies of road safety interventions: searching for needles in a haystack *Journal of Safety Research* 32(3), 267-276

Whitehead A (2002) *Meta-analysis of controlled clinical trials* Chichester: Wiley

Williams AF (1999) Comment on occupant and licensing interventions *American Journal of Preventive Medicine* 16(S1), 6-8

Williams AF and Ferguson SA (2004) Driver education renaissance? *Injury Prevention* 10(1), 4-7

Zaza S, Carande-Kulis VG, Sleet DA, Sosin DM, Elder RW, Shults RA, Dinh-Zarr TB, Nichols JL, Thompson RS and the Task Force on Community Preventive Services (2001) Methods for conducting systematic reviews of the evidence of effectiveness and economic efficiency of interventions to reduce injuries to motor vehicle occupants *American Journal of Preventive Medicine* 21(4S1), 23-30