Risky behaviours: Preferable to crashes for evaluating road safety mass media campaigns?

Wundersitz, L.N. & Hutchinson, T.P.

Centre for Automotive Safety Research, University of Adelaide, Adelaide 5005

Abstract

Decades of research have failed to establish whether or not mass media advertising can reduce road crashes. The probable reason is that the random variability in crash numbers is too great (and, campaigns being very cheap per person reached, even low effectiveness may be enough to be worthwhile). Three alternatives to before-after comparison of crashes as the method of determining effectiveness of an intervention are discussed. These are real-world experiments of high methodological quality, laboratory experiments of the social psychological type, and the measurement of safety-related behaviors. The third of these, before-after comparison of behaviors or variables that can be objectively observed and are closely linked to safety, is suggested as the most promising. However, the behaviors that might plausibly be used as proxies for crashes are quite few in number, and there is an urgent research need to find more of them, together with theory implying that a change in the behavior does indeed mean a change in safety.

Keywords: Media campaigns, Advertising effectiveness, Safety promotion, Road safety campaigns, Accident numbers (variability)

1. Introduction

The question of whether mass media campaigns can improve road safety is controversial. A recent meta-analysis by Phillips et al. (2011) found an average effect of a 9 per cent reduction in accidents. That is large and valuable, if valid. However, many of the studies included by Phillips et al. were methodologically poor (in particular, did not employ randomization), and the average will not convince sceptics, who will say that the average of many weak studies is still a weak result and the picture remains uncertain. Not very different from the issue of mass media campaigns is that of children's education for road safety, and in this context Gillam and Stevenson (1995) were fairly pessimistic, noting (p. 46) the "difficulties inherent in identifying sound pedestrian education programs", and that "methodological flaws, implausible findings and over-confident interpretations abound".

From decades of research of varying quality, a clear result has not emerged. Our view is that this strongly suggests that mass media campaigns do not have a large effect on safety. But, importantly, it remains possible (but not proven) that there is a small saving of crashes and injuries and that advertising is consequently cost-effective. The uncertainty results from the combination of two things. (a) Advertising is very cheap per person reached and hence even a small effect may be sufficient to represent good value for money. (b) There is a desire to measure reductions in crashes, fatalities, and injuries directly, but the variability associated with estimates of crash and injury reductions is sufficiently large that both zero effect and a small effect are compatible with the data.

How might the difficulty in settling the question of cost-effectiveness of publicity, information, and also education and training, be overcome? It might be thought that

employing high quality methodology might permit direct study of crash numbers to achieve this. However, calculations of standard errors and hence necessary sample sizes in section 2 below will establish that it is unrealistic to expect to estimate reductions in crashes, fatalities, and injuries sufficiently accurately --- year-to-year variability in crash numbers is much greater than is generally appreciated. In the hope that some substitute can be found for direct study of crash numbers, sections 3 and 4 will respectively consider (a) laboratory experiments of the social psychological type on changing attitudes, beliefs, and behaviours, and (b) measurement of safety-related behaviours. Section 5 will propose how these potential proxies might be improved, and section 6 is general discussion.

2. Real-world experiments

There is now an influential and articulate body of opinion that strongly favours high quality evidence for public health and social policy decisions. For medicine, see the Cochrane Collaboration, http://www.cochrane.org/, and for social policy, see the Campbell Collaboration, http://www.campbellcollaboration.org/. The concept of high quality evidence encompasses several issues, including the use of valid measures (e.g., being based on what we most care about, such as crashes, rather than precursors such as attitudes and behaviors), having sufficient sample size for reasonably small standard errors, and using good methodology (in particular, experiments employing randomization).

2.1. The principled case for randomization

A controlled experiment employs randomization of the experimental unit (such as a person) to one of two or more different conditions --- for example, treated or not treated (control). In the case of mass media campaigns, it is impracticable to control whether or not a person is exposed to advertisements on television and in the newspapers, but it might be possible to randomly allocate whole cities or states to campaign or no campaign. Randomization has the benefits of avoiding human unfairness and of permitting calculations of the size of chance unfairness. Discussing the disadvantages, in comparison with randomized trials, of observational studies and non-randomized trials, Glonek (2001) wrote as follows:

> "Associations apparent in observational studies and non-randomized experiments cannot be taken at face value as evidence of the true effect of the risk factors in question.... For example, the evaluation of a particular speed enforcement or advertising campaign, or blackspot program, will typically involve the comparison of crash rates before and after the intervention.... In this case it must be expected that many other factors, including the effects of other road safety initiatives and various environmental factors such as the economy, will also influence the crash rates over time. There is no reason to suppose that the joint effect of such other factors should be negligible. In such circumstances, it is not possible to attribute any observed change in the crash rate to the intervention in question."

Before-after comparisons in both treated and control locations are particularly desirable because the general downwards trend in crashes (in developed countries) is not negligible in comparison with the reduction that a modestly successful campaign might achieve.

Researchers and funding bodies who fail to use randomized trials may be scrutinized and criticized. There are valid responses to criticism. Firstly, there is the practicability issue: randomized trials are difficult and expensive to conduct. Secondly, there are some objections of principle to randomized trials. Consequently, while we broadly support those who promote randomized trials, we do not agree with an extreme position. Specific decisions, if taken on

reasonable grounds, are likely to be defensible in the public arena, even if the evidence base is inconclusive or the decision turns out poorly.

The claimed impracticability of randomized trials is particularly plausible in the case of media campaigns, as it is so difficult to control who is exposed to an advertisement and who is not. Presumably, then, the unit of randomization would have to be a whole city (or other geographical area). In itself, this is not impossible. Several cities would need to be enrolled in the trial, and a mass media campaign (e.g., television advertisements) run in a randomly chosen subset of these. Year-to-year changes in road crashes would be monitored in the cities receiving the campaign and in the others, and the two sets of cities compared.

2.2. How many cities would be needed?

Implications of the level of random variability for the sample size required will be derived on the basis of an experiment employing randomization. Of course, the same level of random variability would be present, and the same sample size requirements would apply, in any study using lower quality methods.

In planning a trial, the usual year-to-year variability in crash numbers would need to be known. In calculations below, this will be taken to be 5 per cent of the mean. That is about what it is in South Australia, much greater than would be the case for the Poisson distribution. The size of change in the yearly number of crashes that is of interest also needs to be stated. If the yearly cost of crashes per person is 1000 dollars and the cost of a media campaign per person is 10 dollars (1 per cent of 1000 dollars), a reduction in crashes of 1 per cent would certainly be of interest. Calculations are then made as follows.

- Consider the change in crash numbers in one city from before to after a campaign, and the change in crash numbers in another city from a before period to an after period, but with no campaign there.
- Consider the difference between those changes. There are four city-years of data contributing to this, and its year-to-year variability will be not 5 per cent, but 10 per cent (multiplying by 2, the square root of 4).
- To be “statistically significant”, the difference between the changes will need to be about twice its standard error, i.e., 20 per cent of the average annual crashes in one city.
- The standard error can be reduced by increasing the number of cities. If 400 cities receive the campaign and 400 cities do not, the standard error would be reduced by a factor of 20 (the square root of 400), i.e., to 1 per cent of the average annual crashes in one city.

The conclusion is thus that if it is desired to detect a change of about 1 per cent, it will be necessary to have 800 cities in the experiment. It is possible to argue over the sizes of year-to-year variability and of relevant change (taken here to be respectively 5 per cent and 1 per cent), and four years of data before the intervention and four years of data after would permit the number of cities to be divided by 4, but nevertheless it seems clear this sort of sample size is utterly unrealistic. Having, instead, 8 cities in the experiment might permit a change of about 10 per cent to be detected. Thus an inconclusive result would be quite likely: large enough to imply cost-effectiveness if reproducible, but so small that it is not statistically significant. In addition, such a trial would still not address the possibility of a long-term effect of media campaigns.

People may believe that random variability is much smaller than it really is, and consequently overestimate the chance of getting a clear result based on crashes. When dealing with small numbers of crashes, it is appropriate to assume the number has a Poisson distribution. For example, if the expected number of crashes (e.g., in an area, or to a group of people, or to a fleet of cars) is 10, the standard deviation will be the square root of 10, that is, about 3. But when dealing with large numbers of crashes, as with media campaigns that may be seen or heard by millions of people, this is not appropriate. If the expected number of crashes (e.g., in a city) is 10000, the standard deviation would be the square root of 10000, which is 100, if the Poisson distribution were a valid assumption. But this assumption is grossly wrong: empirically the variation is found to be several times greater (Hutchinson and Mayne, 1977). The reason for the extra variability is poorly understood, but it probably stems from year-to-year variations in such things as the amount of traffic, deliberate interventions in the road and traffic system to make journeys quicker or safer, alcohol consumption, enforcement of the traffic laws, weather, and the reporting of crashes.

3. Laboratory experiments in social psychology

If a real-world trial is likely to be inconclusive, might something more artificial be helpful? Laboratory experiments on effects of advertising certainly can be conducted: the dependent variable might be something like performance in a driving simulator or the participant's self-assessment of how fast they would drive in particular conditions, there can be random allocation of participants to exposure or non-exposure to advertising, and then average performance in the two groups can be compared. An important issue in this line of research is what the theme of the advertisement is, what it is appealing to; fear appeals will be the starting point for the discussion below, followed by appeals to positive emotions. Thus this section is quite narrow in scope, with wider commentary being in section 5.

3.1. Threat and fear appeals

Although a distinction is sometimes made between threat and fear appeals, we will consider them together. Appeals that provoke fear, anxiety or apprehension have been widely used in road safety and other advertising, but after many years of research the effects are far from clear and unequivocal. Essentially, the main steps in the threat appeal technique are to attract and hold the target's attention, generate the fear or anxiety, suggest a safe behavior to cope with the threat, and increase the target's confidence in their ability to successfully and easily perform the safe behavior. The steps relating to efficacy (the third and fourth) are often forgotten by campaign developers (Delhomme et al., 2009). Experimental evidence suggests that threat appeal advertisements that also advise drivers how to drive safely (i.e., effectively avert a threat) are more effective than those that only attempt to stimulate feelings of fear, shock or grief (see Elliott, 2003). For example, Tay and Watson (2002) examined drivers' reactions to fear-based (high threat) fatigue advertisements with and without coping strategies. They found that including coping strategies in high threat messages increased perceived efficacy and consequently increased the likelihood of message acceptance more than fear alone. In addition, the level of fear evoked was correlated with message rejection but not message acceptance leading the authors to suggest that fear could be reduced moderately without reducing message acceptance.

Elliott (2003) examined the literature on fear appeals from 1996 to 2003 and concluded that road safety media campaigns should use fear with caution, as fear arousal can have both facilitating and inhibiting effects and can lead to defective coping mechanisms. Indeed, a

number of studies have found that exposure to fear appeals can elicit maladaptive responses (Schoenbachler and Whittler, 1996; Witte et al., 1998). They are maladaptive in that they do not try to control or remove the threat implied by the fear message, but instead attempt to cope with unpleasant feelings that result. They might include defensively avoiding or ignoring the message, failing to process the threatening part of the message, and denying the personal relevance of the message. Fear appeals may even promote reaction against a message such that individuals view the message as a challenge and increase the undesired behavior.

Recent experimental studies indicate gender might be an important factor that influences the way in which individuals process the relevance of messages using fear appeals (Goldenbeld et al., 2008; Lewis et al., 2008). The threat of physical harm does not appear to be effective for targeting young male drivers, one of the most relevant target groups for which fear appeals were developed (Tay and Ozanne, 2002). Therefore for them other appeals need to be considered. Though reasons for the relevance of gender have been proposed, they seem not to have become widely accepted yet.

In summary, fear appeals can have an impact but only when specific conditions are satisfied. The arousal of fear alone is not enough to adequately motivate behavior. The fear appeal must describe a threat (emphasising the severity of the threat and the vulnerability of the audience), the threat should be personally relevant, and recommendations must be provided for reducing or avoiding the threat (e.g., a safe behavior). Further, the recommendations must satisfy the following points (Donovan et al., 1995).

- Be realistic and credible (i.e., possible to carry out),
- Suggest a specific plan for avoiding the threat (i.e., coping strategy),
- Be perceived as effective and useful to avert the threat (i.e., high response efficacy),
- Allow the target audience to believe that they are capable of carrying out the suggested actions (i.e., high self-efficacy).

Without such recommendations the campaign may be seriously counterproductive as individuals may believe that they are unable to protect themselves from the threat, resulting in defensive and maladaptive responses to the campaign. On this basis, fear appeals should be used with caution and road safety campaign developers should consider using different appeals.

3.2. Appeals using positive emotions

Turning now to alternatives to threat and fear, campaigns can have a rational and/or an emotional approach. Rational appeals provide objective information about the issue and emphasize logic and cognitive processing (e.g., state the benefits of adopting a safe behavior). Emotional appeals emphasize feelings and images and can be positive, negative or a combination of both. Appeals can have elements of both rational and emotional approaches. Emotional appeals can be useful when the target audience already has a strong intention to adopt the safe behavior (Delhomme et al., 2009). Donovan et al. (1995) argue that it is not the type of appeal that is important, but what emotion is relevant to the motivation underlying the decision making for a specific issue. There is a need to obtain a better understanding of the relationship between the extent to which an emotion is evoked and the strength of subsequent attitudinal and behavioral effects.

In contrast to negative emotional or threatening messages that aim to evoke fear, anger, or guilt, positive emotional messages aim to evoke humor, excitement, hope or “good” feelings. Despite calls for more appeals based on positive emotions in road safety, few campaigns have

adopted such an approach. This may be partly due to a lack of knowledge about factors that influence the effectiveness of such appeals, relative to the abundance of literature concentrating on fear appeals. Recent evidence suggests positive emotional approaches can be more effective than negative fear based approaches for males (Lewis et al., 2008a), and there is some political advertising research that suggests positive emotional (or reward) appeals may work better with people who are low on authoritarianism (Wan et al., 2000).

Humor is a positive emotion that has been used occasionally in road safety messages, but there is limited research investigating its effect. Some research has explored the effect of humor in product or commercial advertising, but researchers have questioned whether it can be applied directly to road safety (Delhomme et al., 2009). Evidence from public health suggests that humorous appeals are more persuasive than non-humorous appeals for males in some contexts (Conway and Dubé, 2002; Struckman-Johnson et al., 1994). For speeding behavior, it has been suggested that positive emotional appeals are more effective for males than females, and vice versa for fear-based negative emotional appeals (Lewis et al., 2008a).

There is concern that positive emotional appeals in road safety may be less likely to be recalled and therefore be less effective over time. However, some public health literature suggests that negative appeals might have a diminishing effect over time while positive appeals become more persuasive over time (Lewis et al., 2008b). A recent experimental study exposed participants to road safety advertisements with different appeals and found greater persuasiveness of negative (fear) appeals immediately after exposure but greater improvement for positive humorous appeals over time (up to a month) (Lewis et al., 2008a). Specific expectations regarding the type of message used in road safety might be built up from consistent exposure to a certain type of advertisement. For example, there may be the belief that a humorous appeal is relatively less effective than a fear based appeal because fear is used much more frequently. Lewis et al. (2008b) suggest that the first step towards increasing the effectiveness of positive appeals in road safety campaigns may be to increase their use.

4. Measuring behaviours as an alternative to counting crashes

Laboratory experiments on reactions to advertising seem, then, to be quite a distance from safety on the streets and highways. Returning to real-world research, if (as we have argued) evidence from crashes --- even from randomized trials --- is unlikely to be persuasive, are there any forms of indirect evidence that could stand in its place? An example will now be given of a mass media campaign in which speed was measured rather than crashes. Driving offences and traffic conflicts will then be discussed, as these are among the more attractive of possible proxies for counting crashes.

4.1. Example: Reduction of speed

In the South Australian study of Taylor et al. (2001), an attempt was made to measure the effectiveness of television advertising. The dependent variable was on-road speeds, not the number of crashes. The study lasted three years. Eighteen of the 36 months alternately were designed as advertising months. These began with a three-week period in which there was a high level of advertising or a low level of advertising or no advertising. Speeds were measured in the next week at six locations in Adelaide.

Taylor et al. presented results using several methods of analysis. Details of the results depended on the variable (mean speed, 85th percentile, 95th percentile) and the method of

estimation, but were roughly a 0.2 km/h speed reduction. Again depending on the variable and the method, this sometimes was and sometimes was not statistically significant. Taylor et al. (p. 13) noted that "general perceptions may demand a change of perhaps 5 km/h as being relevant from the point of view of an individual driver". Nevertheless, their view was that the small change observed would imply a "meaningful" and "modest but worthwhile" reduction in casualty crashes.

Each period of advertising lasted three weeks, and there were 12 of them, yet the effects on speed were small in size and of borderline credibility. This study was not designed to measure any long term effects of the advertising, and it gives no evidence one way or the other about whether these exist. It also remains possible, of course, that some future campaign might be more successful than this one was.

4.2. Driving offences and traffic conflicts

Driving offences are a form of behavior, and one for which data is routinely collected (by the law enforcement system). If an intervention (e.g., some form of written publicity or education) can be shown to affect the number of driving offences, is this also evidence that it has an effect on road crashes? It turns out that an effect on offences but not on crashes is quite a common finding in the literature (Peck, 1976; Struckman-Johnson et al., 1989). A possible explanation is a closer linkage between the behaviors targeted by the intervention and being caught offending than between behavior and being involved in a crash. The question remains open whether there is an effect on crashes that is in the same direction as the effect on offences but smaller, or no effect on crashes because the behaviors affected are not relevant to crashes. Kloeden et al. (2008) have suggested that a feasible way forward in research is disaggregation of offences, and disaggregation of crashes: the behaviors in *some types* of offences may be similar to the behaviors in *some types* of crashes.

Traffic conflicts or near misses, at specific sites such as intersections, are sometimes used as a proxy for crashes. Traffic conflicts can be studied by such means as video recording, devices that detect honking horns and screeching tyres, and perhaps pattern recognition methods that operate on digital images. However, traffic conflicts are usually studied at specific locations: they may tell us a lot about the injury crashes that potentially could occur at any individual location. It is not clear how the concept would be adapted to measuring the effectiveness of publicity, the effect of which is intended to be geographically broad. It would presumably be necessary to identify what behavior is targeted by the publicity, create a list of sites where that behavior is likely to occur, and make observations at a sample of these.

5. Improvement of laboratory and real-world research

5.1. Psychological experiments

The majority of fear appeal studies have been conducted in laboratory or experimental settings (Hastings et al., 2004; Witte and Allen, 2000). Hastings et al. (2004) observe that it is these types of studies that have suggested fear appeals can work but such studies have limitations including forced exposure, short-term measurement of effects and an over-dependence on university student samples. The experimental design and contrived setting of these studies means that participants are forced to attend to the advertisements. In a natural setting, individuals may choose not to watch the advertisements after several viewings, or not pay full attention, and consequently the persuasiveness of the message may be reduced or

disappear. There are few real-world evaluations of fear appeals, and their findings usually suggest that fear has a weaker effect (it raises awareness or changes attitudes rather than changes the targeted behavior) and, sometimes, unintended detrimental effects (Hastings et al., 2004). More research conducted in naturalistic settings is needed to overcome these limitations and enable examination of selective exposure, attention, and comprehension. Follow up measures, though difficult, could determine whether the effectiveness of different appeals varies over time.

Another issue with empirical studies designed to evoke emotions, is that they often do not check that the advertisement message is indeed evoking the intended emotion. This is important because different discrete emotions can have different persuasive effects; some inhibit while other facilitate persuasion. If the anticipated emotion is not successfully evoked, then the study is not actually measuring the intended relationship. Rather than merely assuming the intended emotion was evoked, changes in the level of the emotion should be measured via what are termed manipulation checks (Lewis et al., 2009). However, there are problems with self-report measures of individual emotional responses --- for example, possible individual differences in the interpretation of words used to describe emotions (Morris et al., 2002).

Two improvements to the laboratory work may be suggested. (a) Improved understanding of the relation of the psychological concepts under investigation to safety-related behaviors on the road. Understanding the relation of the psychological concepts to the measurements obtainable would also be necessary. It has been proposed that a mental representation of a risk leads to its cognitive evaluation, emotions then follow and lead to action tendencies --- and people are capable of reporting their action tendencies (Böhm and Pfister, 2000; Xie et al., 2011). (b) A greater concentration on behaviors rather than feelings in the experiments --- for example, speeds chosen in driving simulators and gap acceptance in virtual environments have already been used (e.g., Algie et al., 2008). Technological improvement in simulated environments are continually taking place, and this area holds some promise.

5.2. The link between behavior and safety

We want to be able to infer safety from behavior. Consider the three classes of behaviors or observable variables below.

- Examples of behaviors or variables that are closely linked to safety include: blood alcohol concentration (and breath alcohol concentration), speed, and the usage of secondary safety devices (seat belts, child restraints, motorcycle helmets, cycle helmets). In each case, a change would be seen as implying a change in safety. An important limitation would be that quantitatively it would not be known how much change in safety would result from a given observed behavioral change.
- Further examples of behaviors that are observable are as follows: driver head movements at junctions, pedestrians crossing at designated crossings rather than elsewhere, traffic conflicts, taking a rest break on long journeys, gap acceptance, various driving offences. Our view, however, is that their link with safety is not sufficiently tight that a change in these behaviors would be seen as necessarily implying a change in safety.
- There are scales and questionnaires that purport to measure attitudes and self-reported behaviors. This type of evidence will not usually persuade those who are sceptical about whether mass media campaigns are effective. Firstly, there are substantial measurement difficulties, including limited reliability and validity of these scales.

Secondly, the connection between a change in attitudes or self-reported behavior and crash numbers is insufficiently well established.

The plausibility of the link between behavior and safety is vital. It is not clear, however, on what basis plausibility is judged. For the first class of behaviors above, the logic is obvious. In the other cases, we suggest some form of reasoned theory is highly desirable. The theory might, for example, involve perception of risk leading to change of behavior. Brown (2010) reports on a three-observation design in which a media campaign was conducted between Time1 and Time2, and changes from Time1 to Time2 in risk perception were found to be predictive of changes from Time2 to Time3 in self-reported behavior. Obviously, as Brown concedes, one would ideally also want to strengthen the link between self-reported and objectively observed behavior. (It so happens that, like that of Taylor et al., this was a South Australian study of speeding.)

The more detailed the theory connecting behavior and safety, the better can its plausibility be judged. Speed is probably the most broadly plausible indicator of danger or safety: reduce speed, and a reduction of crashes is very likely. That argument has some force on its own, but it is improved if details are included about the speed range and circumstances in which a lot of crashes happen. If a countermeasure is shown to reduce speeds in the speed range within which most crashes occur, there can be confidence that the overall risk of crashing is reduced appreciably, but if the effect on speeds is confined to a speed range in which crashes are few, its effect on crashes will be small. Similarly, a reduction of speed among the types of drivers who have most crashes or at the types of site where crashes are most frequent will suggest a worthwhile reduction of the overall risk of crashing, but an effect on safe drivers or at safe sites will suggest only a small effect on crashes.

5.3. A theory of theories?

Suppose it is claimed that some theory --- or perhaps common sense --- tells us that so-and-so change in behaviour will lead to a reduction in such-and-such type of accident. How can we judge the correctness or plausibility of the one thing leading to the other? It seems that we need some theory about theories.

Archer (2005) concentrates on what are termed proximal safety indicators (very closely related to crashes, e.g., traffic conflicts and measurements related to the conflicts), and also pays some attention to safety-influencing factors (these are less immediately related, e.g., speed). He refers to the events being "representative of the same underlying processes" as crashes, having a "statistical and causal" relationship with crashes, and to a "common causation process" (pp. IX, 42, 46). If objective and widely-accepted criteria existed for deciding whether an indirect measure can substitute for a count of crashes, Archer would have referred to them; we take the phrases quoted to be evidence that it is unfortunately necessary to rely on common sense, or, if we can spell out details step by step, theory.

Our view is that the link between behavior on the road (or performance in the laboratory) and crash occurrence is typically one whose strength, and even existence, is uncertain, and that much more needs to be done first to articulate, second to validate or disprove common sense ideas, and third to quantify the connection. Unfortunately, a satisfactory theory of theories is probably decades away. It may be suggested, though, that there is likely to be confidence in a theory if it is explicit, with many steps being spelt out, each having been tested and supported.

5.4. Two comments on statistical testing

Is conventional statistical testing, using an α level of .05, appropriate? It might be felt it is very valuable to detect a reduction in crashes when one has in fact occurred (Parkhurst, 1990), and that consequently an α level of perhaps .20 should be used. We have some sympathy with this view. Nevertheless, it is unlikely that anyone will be satisfied if a large standard error attaches to a change in crash numbers: it will continue to be of interest to find behaviors that permit statistical tests of higher power and for which a close link to crashes is credible.

Could it even be that casting the discussion in section 2.2 in terms of statistical significance was responsible for our claim that hopes of using crash numbers to directly evaluate media campaigns are unrealistic? Perhaps analysis of cost effectiveness (without testing statistically for the existence of an effect) would be more appropriate? That is unlikely, as rather than leading to a demand for an enormous sample size, the random variability would appear in inconsistency of results, with campaigns being apparently cost effective at some times and places and not cost effective at other times and places.

6. Discussion

The evaluation of mass media campaigns is clouded by the question of whether a real reduction in crashes can be detected. (a) If the large standard error of the estimated effect size is appreciated only after the study has been conducted, the lack of a statistically significant effect may properly be judged to be not compelling. (b) If, on the other hand, the low statistical power is appreciated at the planning stage, this may prevent research being done that is largely pointless (assuming that statistical tests will be used in the conventional way). (c) Or a higher-than-usual α level might be employed in statistical tests. (d) It is sometimes the case that, in whatever specific situation is of interest, many different things are causing crashes. Then a real effect operating on only one of these may be virtually undetectable with ordinary research methods. In this case, efforts could be made to specify, theoretically or otherwise, the class of crashes that will be affected and then identify these in the dataset.

In this paper, we have concentrated on stepping further away from usual practice.

- Retaining crash numbers as the dependent variable of interest, the methodology of research could be improved by conducting real-world experiments. This is unlikely to be a solution, however.
- Greater use could be made of laboratory experiments in establishing the psychological foundations of the effects of mass media. Unfortunately, what is measured is sufficiently far from real-world safety that it is uncertain whether any effects detected would also apply to crashes.
- Greater use could be made of proxies for crashes, i.e., behaviors or variables that are closely linked to safety and can be objectively observed.

Better theory is needed in order to link any results of the second and third types (i.e., indirect methods) with crashes, and to indicate how generalizable are results obtained with direct methods. Hoekstra and Wegman (2011) made a distinction between rational, automatic, mimicking, and socially conforming behaviors, and this carries a suggestion of theory.

There are good reasons why there is still doubt about whether or not road safety mass media campaigns have a small but worthwhile effect: it is unreasonable to expect crash numbers to eventually answer the question, as the year-to-year variability in crash numbers is too great. This may apply to risks other than road crashes and countermeasures other than media

campaigns. There are two key features: (a) the countermeasure is cheap and applies to a lot of people, so that it may be worthwhile even at low effectiveness, and (b) the risk is a count of events and observed at an aggregate level, so that one needs to estimate the variability empirically, rather than rely on variability being purely that of the Poisson process.

We are fairly optimistic about the measurement of safety-related behaviors or variables; indeed, this already occurs, as discussed in section 4. However, the behaviors that might plausibly be used as proxies are quite few in number, and there is an urgent research need to find more of them --- along with theory that will persuade us that a change in the behavior will indeed reflect a change in safety. We principally have in mind behaviors that are not recorded and would need observing specially. In contrast, a behavior that is recorded administratively is commission of driving offences. As we have discussed, these may be affected by interventions differently from crashes, but it is possible that some types of offences may be similar to some types of crashes, and that disaggregations of results offer another way forward.

Acknowledgements

The Centre for Automotive Safety Research, University of Adelaide, is supported by both the South Australian Department of Planning, Transport and Infrastructure and the South Australian Motor Accident Commission. The views expressed in this paper are those of the authors, and do not necessarily represent those of the University of Adelaide or the sponsoring organisations.

References

Algie, J., Jones, S. & Rossiter, J.R. 2008. Examining the effect of fear patterns with repeated exposures. Proceedings of the World Social Marketing Conference, http://www.docin.com/p-65995468.html; pp. 124-132.

Archer, J. 2005. Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: A study of urban and suburban intersections, Doctoral thesis, Centre for Traffic Research, Department of Infrastructure, Royal Institute of Technology, Sweden.

Böhm, G. & Pfister, H.-R. 2000. Action tendencies and characteristics of environmental risks. *Acta Psychologica*, 104: 317-337.

Brown, S.L. 2010. Changes in risk perceptions prospectively predict changes in self-reported speeding. *Risk Analysis*, 30(7): 1092-1098.

Conway, M. & Dubé, L. 2002. Humor in persuasion on threatening topics: Effectiveness is a function of audience sex role orientation. *Personality and Social Psychology Bulletin*, 28(7): 863-873.

Delhomme, P., De Dobbeleer, W., Forward, S & Simões A. (editors) 2009. Manual for Designing, Implementing and Evaluating Road Safety Communication Campaigns. Brussels: Belgian Road Safety Institute.

Donovan, R., Henley, N., Jalleh, G. & Slater, C. 1995. Road Safety Advertising: An Empirical Study and Literature Review. Canberra: Federal Office of Road Safety.

Elliott, B.J. 2003. The psychology of fear appeals re-visited. Paper presented at the Road Safety Research, Policing and Education Conference http://www.rsconference.com/pdf/RS030056.pdf

Gillam, C. & Stevenson, M. 1995. Evaluation of pedestrian safety education programs for young children: A review of methodologies. *Asia-Pacific Journal of Public Health*, 8: 42-47.

Glonek, G. 2001. On the strengths and weaknesses of experimental and observational methodologies. In: Road Safety Research. Proceedings of the National Speed and Road Safety Conference, held in Adelaide.

Goldenbeld, C., Twisk, D. & Houwing, S. 2008. Effects of persuasive communication and group discussions on acceptability of anti-speeding policies for male and female drivers. *Transportation Research*, *Part F*, 11: 207-220.

Hastings, G., Stead, M. & Webb, J. 2004. Fear appeals in social marketing: Strategic and ethical reasons for concern. *Psychology and Marketing*, 21(11): 961-986.

Hoekstra, T. & Wegman, F. 2011. Improving the effectiveness of road safety campaigns: Current and new practices. *IATSS Research*, doi: 10.1016/j.iatssr.2011.01.003

Hutchinson, T.P. & Mayne, A.J. 1977. The year-to-year variability in the numbers of road accidents. *Traffic Engineering and Control*, 18: 432-433.

Kloeden, C.N., Anderson, R.W.G. & Hutchinson, T.P. 2008. Differences between groups of drivers: Offences contrasted with crashes. *Flinders Journal of Law Reform*, 10(3): 815-830.

Lewis, I., Watson, B. & White, K. 2008a. Predicting future speeding behaviour: The appeal of positive emotional appeals for high risk road users. Paper presented at the High Risk Road Users Conference, Joint ACRS-Travelsafe National Conference, Brisbane.

Lewis, I., Watson, B. & White, K. 2008b. An examination of message-relevant affect in road safety messages: Should road safety advertisements aim to make us feel good or bad? *Transportation Research*, *Part F*, 11:403-417.

Lewis, I., Watson, B. & White, K.M. 2009. What do we really know about designing and evaluating road safety advertising?: Current knowledge and future challenges. In: R. Grzebieta & D. McTiernan, editors. Proceedings of the 2009 Australasian Road Safety Research, Policing and Education Conference and the 2009 Intelligent Speed Adaptation Conference. Sydney: Roads and Traffic Authority of New South Wales (ISBN of USB: 978-1-921692-26-0); pp. 733-746.

Morris, J.D., Woo, C., Geason, J.A. & Kim, J. 2002. The power of affect: Predicting intention. *Journal of Advertising Research*, 42: 7-17.

Parkhurst, D.F. 1990. Statistical hypothesis tests and statistical power in pure and applied science. In: G.M. von Furstenberg (editor). Acting Under Uncertainty: Multidisciplinary Conceptions. Norwell, MA: Kluwer; pp. 181-201.

Peck, R.C. 1976. Toward a dynamic system of driver improvement program evaluation. *Human Factors*, 18(5): 493-506.

Phillips, R.O., Ulleberg, P. & Vaa, T. 2011. Meta-analysis of the effect of road safety campaigns on accidents. *Accident Analysis and Prevention*, doi: 10.1016/j.aap.2011.01.002

Schoenbachler, D.D. & Whittler, T.E. 1996. Adolescent processing of social and physical threat communications. *Journal of Advertising*, 25: 37-54.

Struckman-Johnson, C., Struckman-Johnson, D., Gilliland, R.C. & Ausman, A. 1994. Effect of persuasive appeals in AIDS, PSAs and condom commercials on intentions to use condoms. *Journal of Applied Social Psychology*, 24:2223-2244.

Struckman-Johnson, D.L., Lund, A.K., Williams, A.F. & Osborne, D.W. 1989. Comparative effects of driver improvement programs on crashes and violations. *Accident Analysis and Prevention*, 21(3): 203-215.

Tay, R. & Ozanne, L. 2002. Who are we scaring with high fear road safety campaigns? *Asia Pacific Journal of Transport*, 4:1-12.

Tay, R. & Watson, B. 2002. Changing drivers' intentions and behaviours using fear-based driver fatigue advertisements. *Health Marketing Quarterly*, 19(4): 55-68.

Taylor, M.A.P., Woolley, J.E. & Dyson, C.B. 2001. Outcomes from the South Australian road safety media evaluation study. In: Road Safety Research. Proceedings of the National Speed and Road Safety Conference, held in Adelaide.

Wan, F., Meirick, P., Williams, J., Holmes, J. & Feibich, C. 2000. Threat, authoritarianism and political advertising: An experiment in personality and persuasion. Paper presented at the Association for Education in Journalism and Mass Communication (AEJMC), Phoenix, AZ.

Witte, K. & Allen, M. 2000. A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health Education & Behavior*, 27: 591-615.

Witte, K., Berkowitz, J.M., Cameron, K.A. & McKeon, J.K. 1998. Preventing the spread of genital warts: Using fear appeals to promote self-protective behaviors. *Health Education & Behavior*, 25: 571-585.

Xie, X.-F., Wang, M., Zhang, R.-G., Li, J. & Yu, Q.-Y. 2011. The role of emotions in risk communication. *Risk Analysis*, doi: 10.1111/j.1539-6924.2010.01530.x