Same Data, Different Conclusions:
Analysis of the New Zealand Drink-Driving Campaign Data

Every Picture tells a Different Story

Tony Lewis

Everybody has an opinion on whether the New Zealand road safety television advertisements work to persuade people to behave better and thus reduce the road toll. The TV viewers disagree in conversation, and the experts disagree with one another in written reports, but the public wants to know whether the money is well spent and they can't understand why the experts can't tell them. This paper explains, in a language that is accessible to everyone, why we are so bad at monitoring the effects of advertising campaigns such as the road safety campaign. The paper explains how different results can be obtained from the same set of data and suggests that there is no objective way of judging between the different results. Moreover, the paper makes the claim that no amount of expertise can achieve a reliable result. The paper claims that the problems are inherent in the way the data are generated and collected, but a makes a controversial suggestion for a change to the way we view social experiments, so that the data generated is amenable to reliable analysis.

Keywords: drink driving statistics, advertising campaign effectiveness, road safety

Introduction

In Social Marketing, as elsewhere in Marketing, reliable measurement of advertising effectiveness continues to be elusive. The aim should be to predict how an advertising campaign will change behaviour, but the usual measures have to do with how many people see, like and remember the advertisements. The task of measurement is made more difficult by the fact that the data are usually non-experimental, with no controls on how it is produced.

Data emanating from the New Zealand drink-drive advertising campaign, which began in earnest in October 1995, have been analysed for effectiveness at least seven times. Gregg (1996) and Falconer (1996) found the campaign to be successful in reducing the road toll. Macpherson and Lewis (1996) could not find any evidence of an improvement in drink-driving behaviour as a result of the campaign. Bliss, Guria, Vulcan and Cameron (1998), and Cameron and Vulcan (1998) found it to be successful, as did Tay (1999). White (2000) could not find any relationship between the campaign and serious accidents, and criticised the results of Bliss et al. (1998) and of Tay (1999).[1]

How can this be, and what are the public and politicians to make of it? What can the public do when a bunch of highly paid specialists in data analysis, with an impressive list of qualifications between them, tell conflicting stories about the use of advertising in road

---

[1] In the rest of the paper, only the first author's name will be given in the references.

safety campaigns? Who are the public to believe; the last expert to do the analysis or the first, or the ones in between? It is no use going to the publications, they are written in an arcane language, that even other experts sometimes struggle to comprehend; the public has no show.

There is plenty of criticism by analysts of other analysts who have come up with a different result from their own. The criticism is usually about technical aspects of the analysis. The other authors did something wrong, they didn't take account of this or that possible effect, their conclusions are not supported by their results. The criticisms are written in the same arcane language that is used in the reports. All this is of no interest to policy makers and the public, who cannot understand the criticisms anyway. "Just get on with the job", they say, "that is what we pay you for."

This paper tells the story through pictures, to show how conflicting results can be obtained by different analyses of the same data, so that anyone can understand the problems, and makes a suggestion about how data can be collected in the future to make easier the task of monitoring any efforts to reduce dangerous driving behaviour. There is some argument about whether we should be looking at driving behaviour, or the results of bad behaviour, to measure the effect of a campaign. That is, some analysts think we should be measuring the effect by seeing if drivers drink less, other analysts think we should be measuring the effect by seeing if the accident statistics improve. The lead of Macpherson is followed here and this paper uses drink diving behaviour as measured by the number of positive breath tests, but that is just a detail, and does not alter the fundamental problem that will be explained with pictures.

Main Explanation

Figure 1 shows the number of positive breath tests, called evidential breath tests, prior to and following the TV advertising campaign which began in October 1995.

The first thing to notice is that the positive tests generally peak in December, perhaps people live it up a bit in the holiday season, and drop off a bit in the middle of the year.

We can measure this effect, insofar as it occurs regularly, and remove it by deseasonalising the data. It is the sort of thing the Statistics Department does when they try to abstract from the effect of school leavers coming on to the job market or from the effect or the fruit picking season, when they measure unemployment.

Figure 2 shows the movement of positive tests when these regular seasonal effects are removed.

Here we see a distinct drop in the number of positive tests after mid 1995. Gregg and Falconer were relying on this observation for the conclusion that the campaign was effective. They actually used fatal crashes and serious crashes to measure the effect, but the same pattern applies to those data.

Actually, though the graphic images of the results of traffic accidents were screened from October 1995 on, more conventional advertising had been taking place for some time. We can measure the public exposure that a series of advertisement gets by a number called Target

Audience Rating Points (TARPS), obtained from one of the TV rating companies. These numbers are made up by estimating the number of people who watch particular programmes at the time the advertisements were run. Advertisers and 'I'V companies use TARPS to help them schedule advertisements and to charge out air-time.
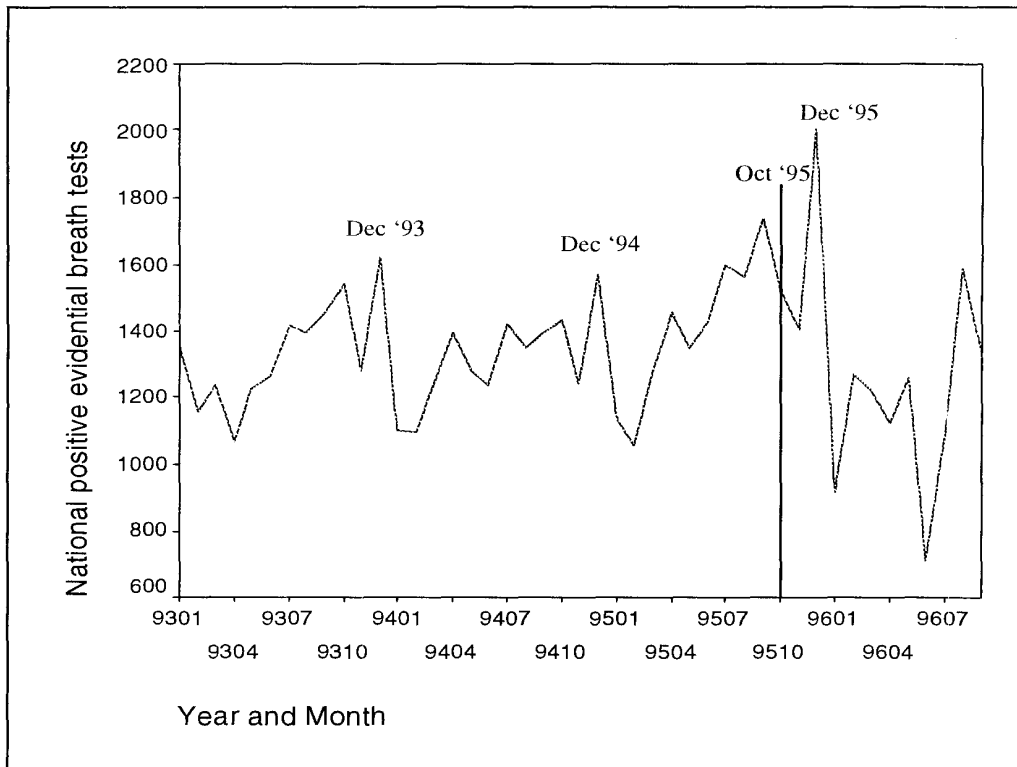

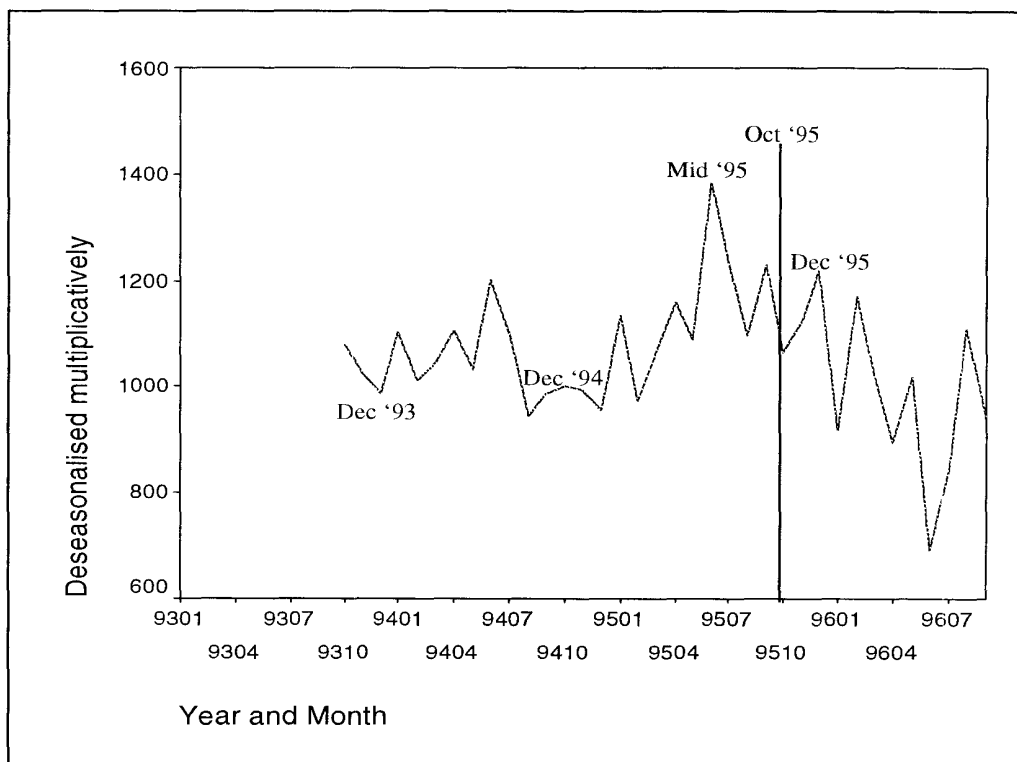**Figure 1. Time series of evidential breath tests (EBTs)**

Figure 2. Time series of de-seasonalised EBTs

Macpherson made an adjustment to these figures to take account of the fact that an advertisement might have a lingering effect. These adjusted numbers are called in the trade, "Adstock".

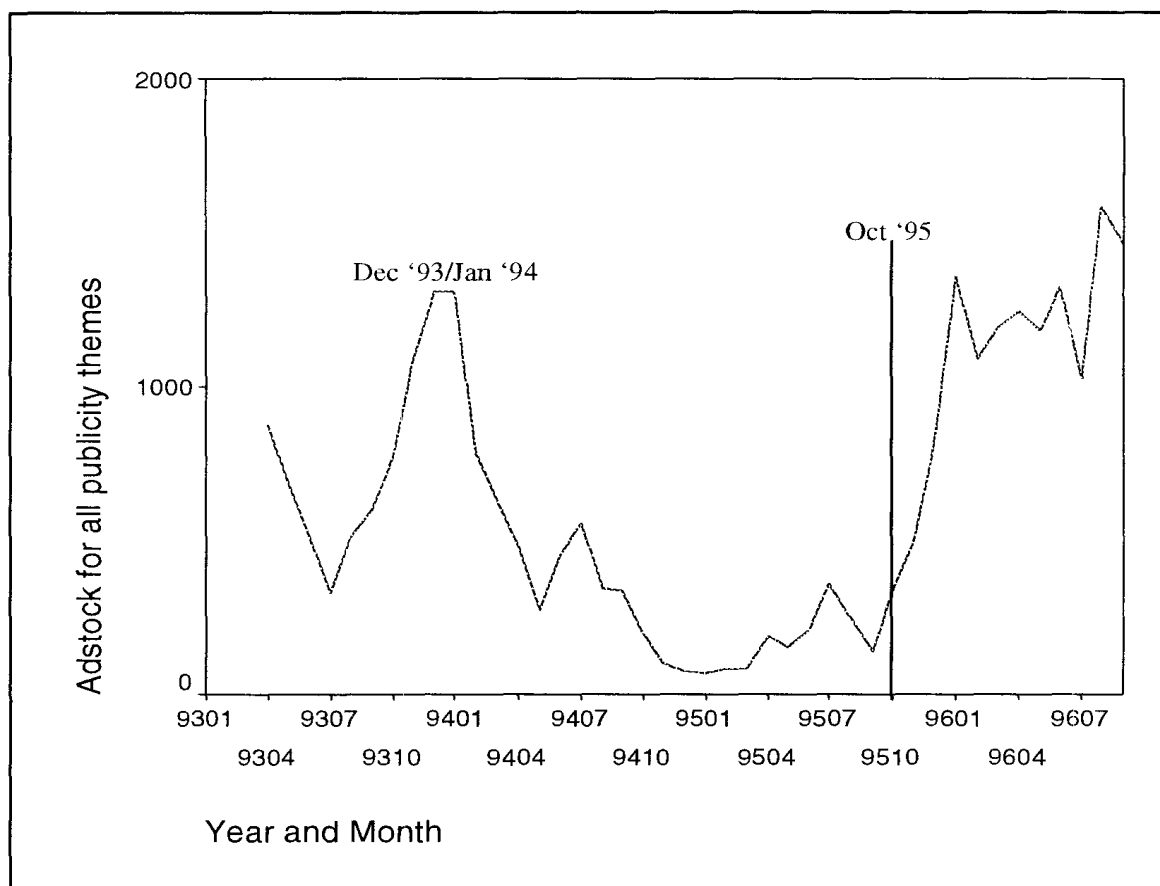Figure 3 shows the movement of Adstock over about the same period as the EBTs were plotted.

Figure 3.   Time series of adstock for all road safety advertisements

There was a drop in advertising from the beginning of 1994 to late 1995, however, referring back to Figure 2, we don't notice any effect of that drop on EBT numbers over that time. But look at the drop in EBTs shown in Figure 2 after the blood and guts advertising began in late 1995. It looks as though Falconer Gregg, Bliss, Cameron and Tay were right, and Macpherson and White were wrong.

Just to check, we should see if there were more tests carried during this period, because if more tests are being made, presumably more people will be caught with excess alcohol on their breath. The tests you have to take if you are stopped by a patrol are called Compulsory Breath Tests (CBTs).

Figure 4 below shows the movement of Compulsory Breath Tests over the period.
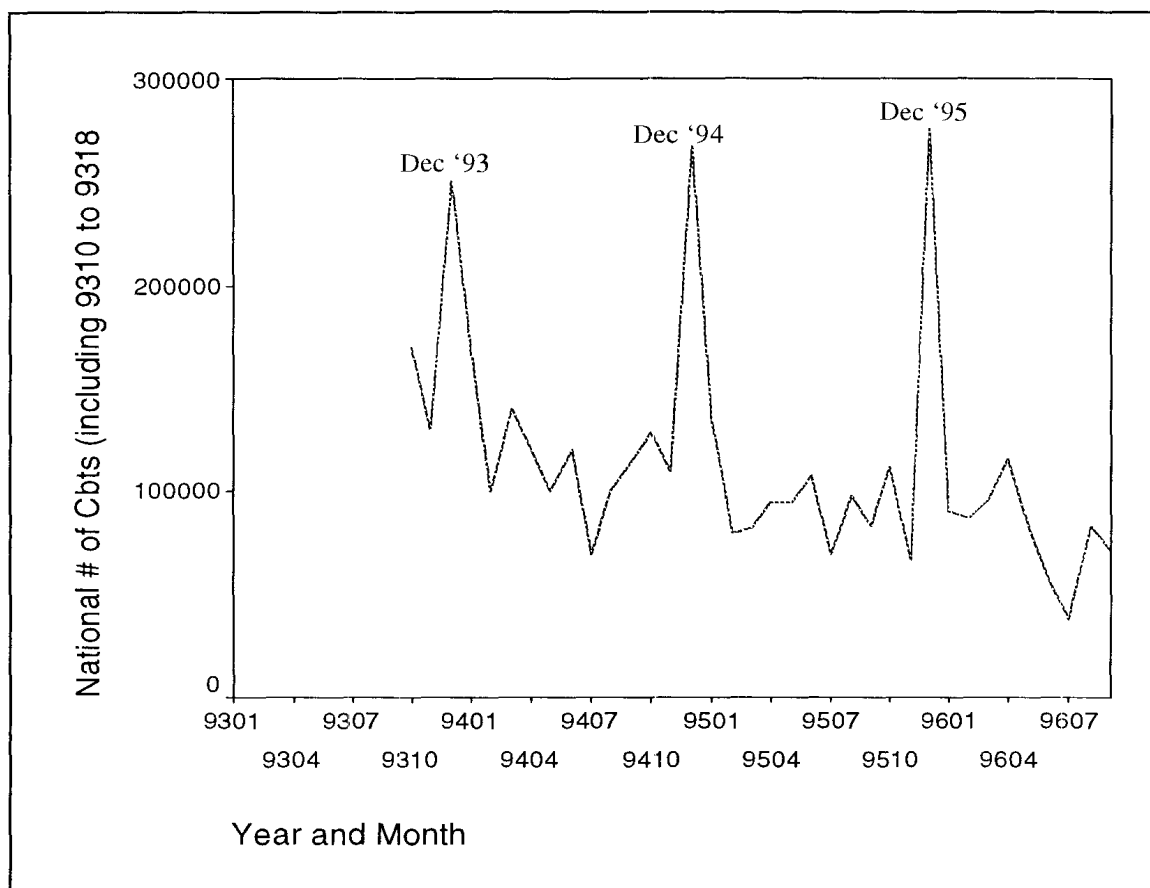
Figure 4.   Time series of compulsory breath tests (CBTs)

Certainly more tests are carried out over the holiday period, which might account for the end-of-year peaks we noticed earlier. Perhaps we should see if taking this into account will alter our results. We can do this by calculating the number of positive tests, per test carried out; we will call this number the ratio of EBTs to CBTs. If, for instance, every 1 in 100 breath tests was positive, our ratio would be 0.01 (M). It' this number jumped to 0.02 (2%) this would mean people were behaving worse; in proportion to the tests being carried out, twice as many people were being caught. If it declined to 0.005 (t/z%), this would mean people were behaving better; in proportion to the tests being carried out, only half as many people were being caught.

Figure 5 shows the movement of the ratio of positive tests, to tests carried out.

Now it appears that since the campaign began in October 1995, after an improvement to December 1995, people have been behaving worse. In fact there seems to be a trend going upwards from the beginning of 1993, which is the earliest date for which we have data. It looks as though Falconer, Gregg, Bliss, Cameron and Tay were wrong after all, and so was Macpherson. Macpherson could not find a relationship between the level of advertising and the drink-driving behaviour, but Figure 5 shows there is one; the only problem is that it was a positive one and EBTs increased to June 1996!

Actually, over this period the number road fatalities dropped, as shown in Figure 6
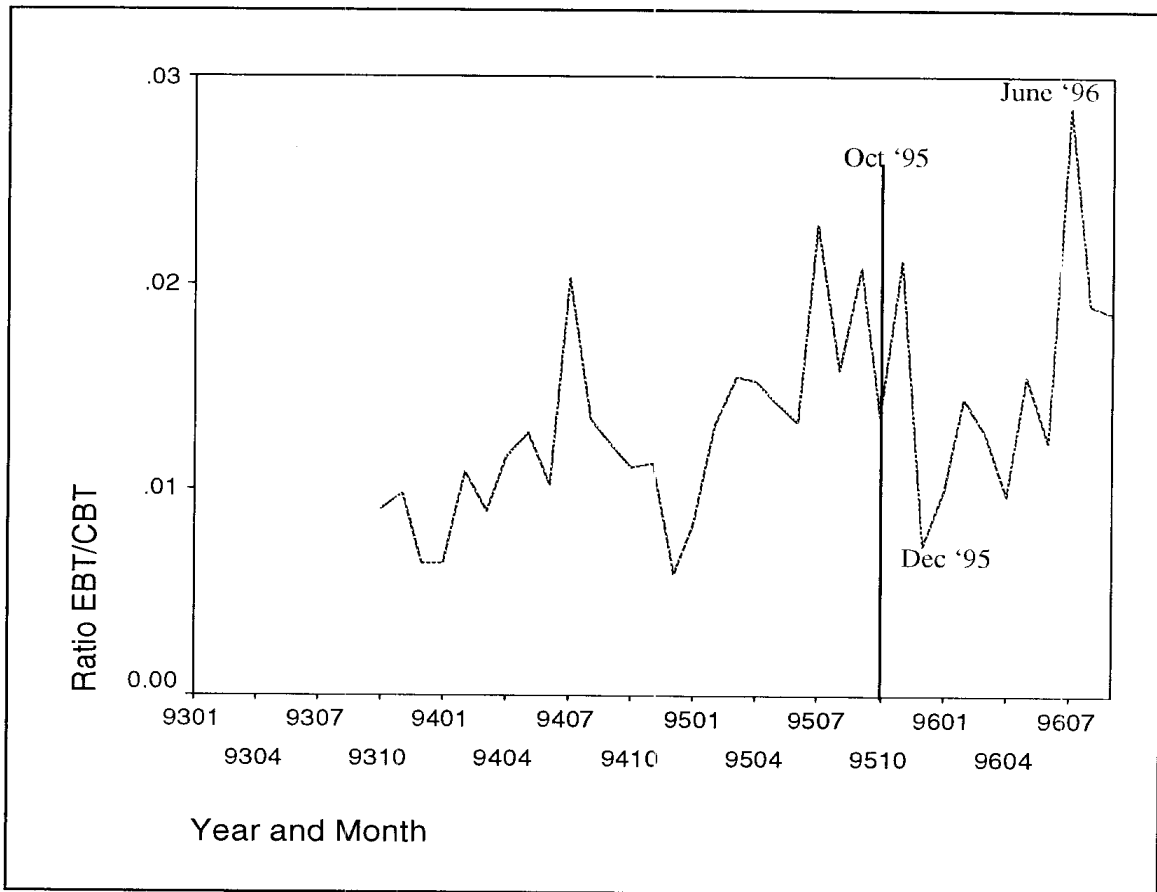
Figure 5. Time series of ratio of EBS to CBTs (EBTs/CBTs)
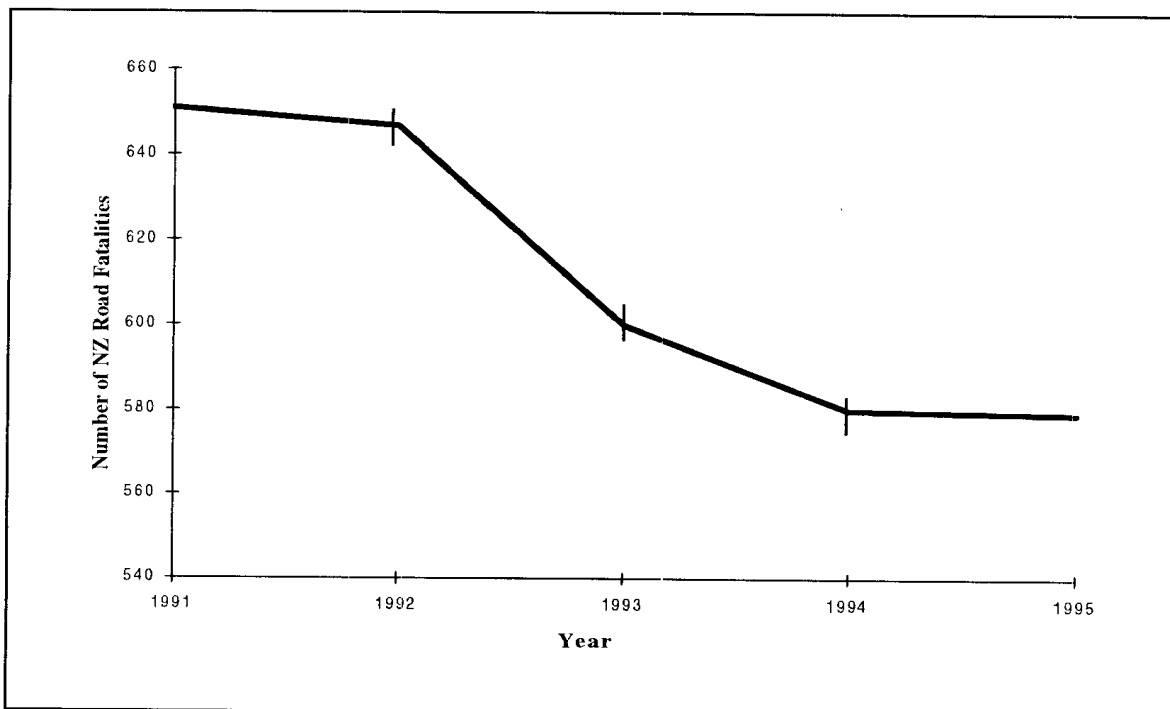

Figure 6. New Zealand road fatalities (1991 to 1995)

Source: Land Transport and Safety Authority (Sept 1996). *Road Deaths.* Official road fatality statistics prepared by LTSA Research and Statistics section.

This is the graph that Macpherson was relying on when he claimed that the drop in road fatalities that Falconer and Gregg used to support their case that the campaign was an effective means of reducing the road toll, was just part of a continuing trend. White used data similar[2] to this to show that a downward trend and new car registrations could explain most of the variation in serious casualty crashes.

White used new car registrations and unemployment as proxies for economic conditions in his analysis. Somebody investigating the safety of new cars as opposed to older cars, might have come to the conclusion that new cars are safer than old cars, and have not given a thought to the drink driving, advertising campaign!

We ought to see what the ratio plot looks like when we de-seasonalise these data in the way we did for the EBTs, just in case the results change again. Figure 7 shows this de-seasonalised ratio.
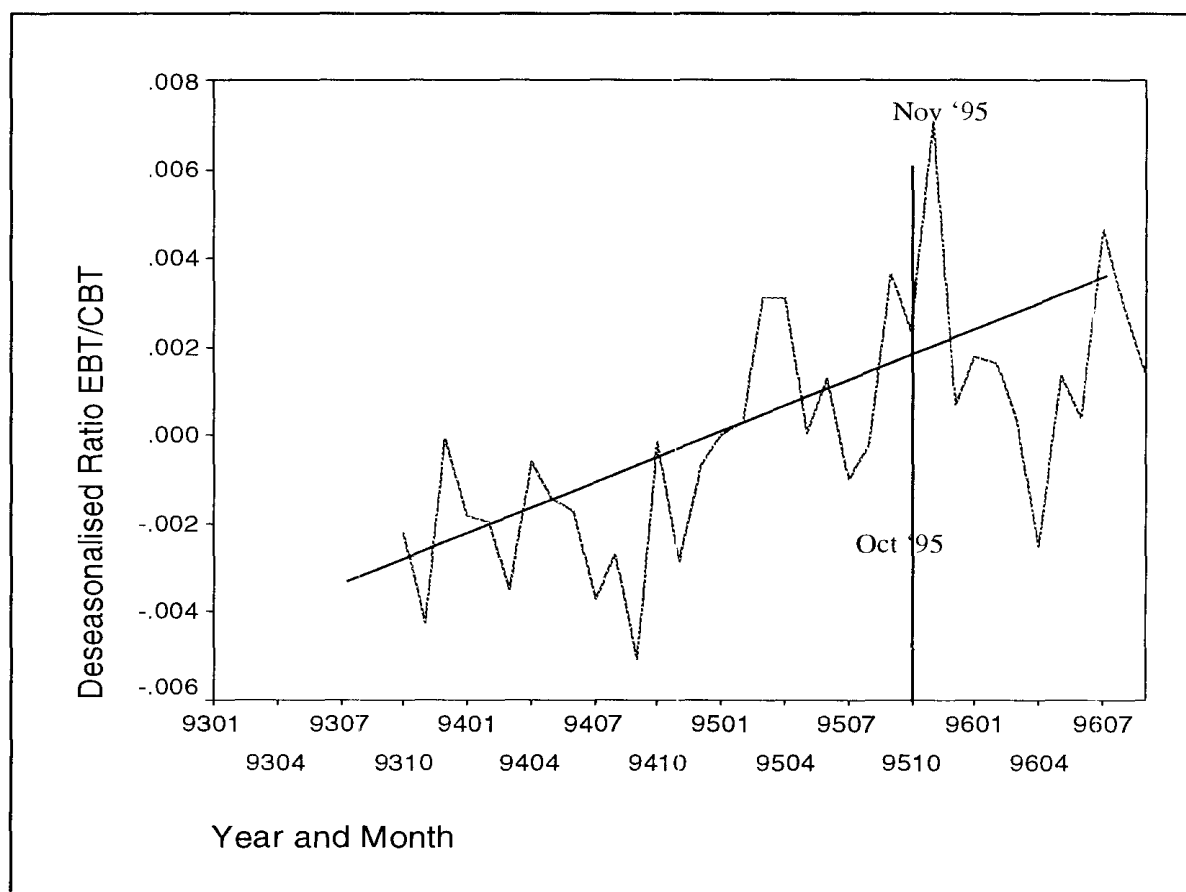

Figure 7.   Time series of de-seasonalised ratio EBTS to CBTs
January 1995 has been used as a base for deseasonalising

It makes no difference, people have been behaving worse over the whole period, though this may have leveled out since the campaign began in October 1995. It is apparent that there is something of a trend towards worse behaviour over the period since late 1993, as shown by the upward sloping line. There is no accounting for the increase in November 1995, just after the campaign began.

---

2 White used serious casualty crashes data as an indicator of the effects of the campaign, instead of TBTs.

Actually, if we take out the upward trend, then you can see that Macpherson was right after all; there does not seem to be any relationship between the advertisements and drink-driving behaviour. You can see this by comparing the scatter around the line, with Figure 3 showing Adstock; clearly there is no relationship. Adstock is high at both the beginning and end of the period, but EBTs just seem to be erratic.

However, instead of taking the ratio of EBTs to CBTs to allow for increases in enforcement, we could try to account for the effect of the change in visible enforcement. To do this we estimate the effect of a 1 % increase in CBTs on EBTs, and this gives us a different story again. Macpherson calculated this to be about 0.4%. This means that if the number of tests is doubled (increased by 100%), EBTs are not doubled as one would expect, but would just increase by 40%. It could be claimed that this number being so much less than expected, shows that the greater enforcement presence (more police on the roads making tests) induced drivers to behave better. But it could just be that the worst trouble spots are covered on a routine basis, and as the number checkpoint stations increases, the police have to check places that are less of a problem.

Figure 8 shows the time series of the EBTs if the monthly variation and the CBTs were the only reliable component of the EBTs time series. They are the EBT values if we regard the monthly variation and the variation of the CBTs as the sole cause of systematic variation in EBTs.
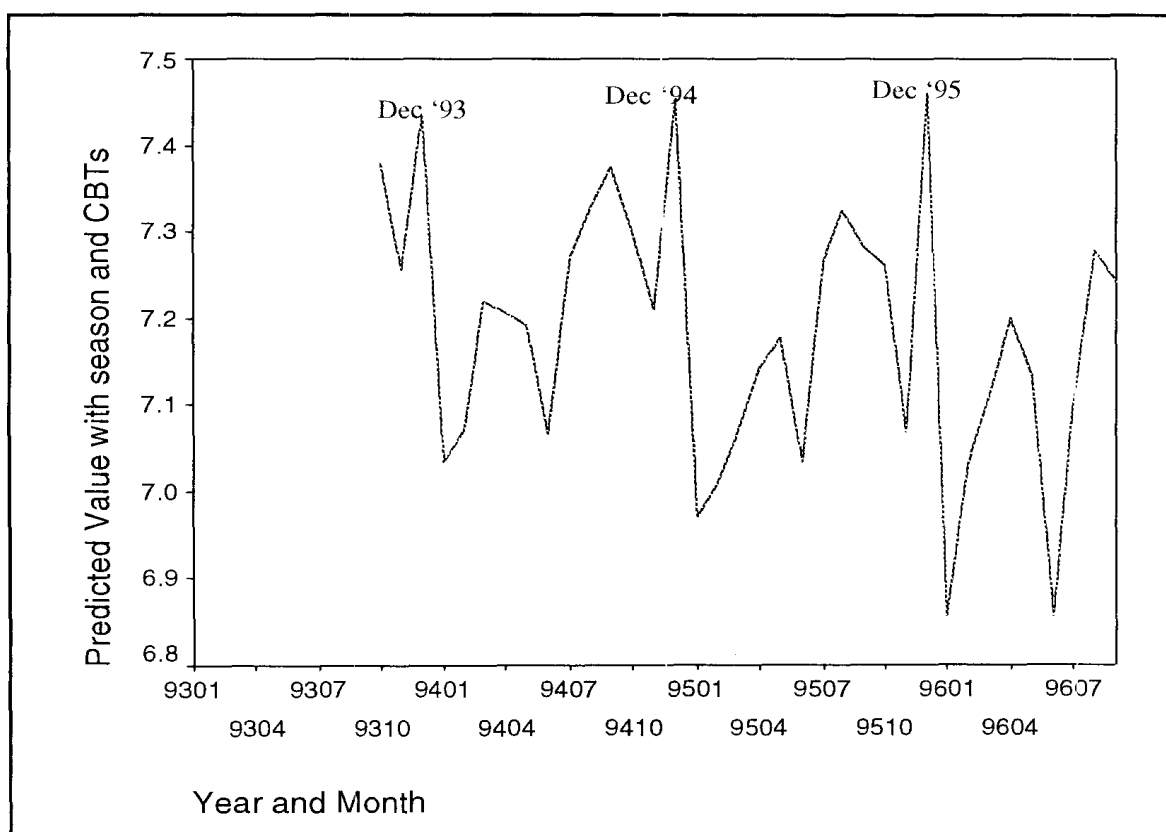

Figure 8.   Time series of predicted values of EBTs using CBTs and monthly factors
Note: The EBT values on the vertical axis are scaled (logarithmically) to aid interpretation

Once again the apparent dramatic increase in bad behaviour in December shows up, but we now know that it is the extra December CBTs that are causing the December peaks in EBTs.

If we calculate the monthly effect after first removing the effect of the variation in CBTs, we would get a new picture.

Taking January as a base because it is generally a low month in the "de-CBTised" (for want of a better word) series, we can calculate the ratio of the other monthly averages, to the January average. The numbers are given in Table 1 below.

Table 1. Monthly factors

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
A	1	1.09	1.20	1.26	1.24	1.03	1.29	1.43	1.41	1.43	1.25	1.64
B	1	1.20	1.27	1.31	1.36	1.14	1.62	1.56	1.56	1.41	1.35	1.34

A gives the monthly factors (the scaling factors ictured in Figure 8) calculated from the EBTs.
B gives monthly factors calculated if CBTs are taken into account.

Now we can see why Macpherson got the result he did. According to Figure 3, Table 1 shows how we can be deceived when analysing a data series without giving sufficient thought to the other things that might be of influence. Using the original series as given in Figure 1 (line A in Table 1), it appears that January is the best month in terms of driver behaviour, and June is not far behind, but EBTs are 64% greater in December than they are in January. Using the de-CBTised series (line B in Table 1), January is still the best month, but the worst month is now shown to be July, which is 62% greater than January, with August and September not much better.

Finally we can de-seasonalise and "de-CBTise" the data (we might call this data "corrected" for monthly influences and for changes in CBTs) and draw the same conclusion as that of Macpherson. This time series is given in Figure 9.

Adstock rose in late 1993, then fell away to nothing in late December 1994 and early January 1995, but "corrected" EBTs just continued to fluctuate as shown in Figure 9. When the campaign began in October 1995, EBTs rose sharply to a peak in December 1996 as the campaign was gathering momentum, but then dropped away again; remember that December is not actually a peak month, if the series is "corrected" for changes in CBTs. Perhaps the reason for Macpherson's results was that the period of apparently random fluctuation up till the beginning of the campaign and the December 1996 figure, swamped the fall in EBTs after December 1996.
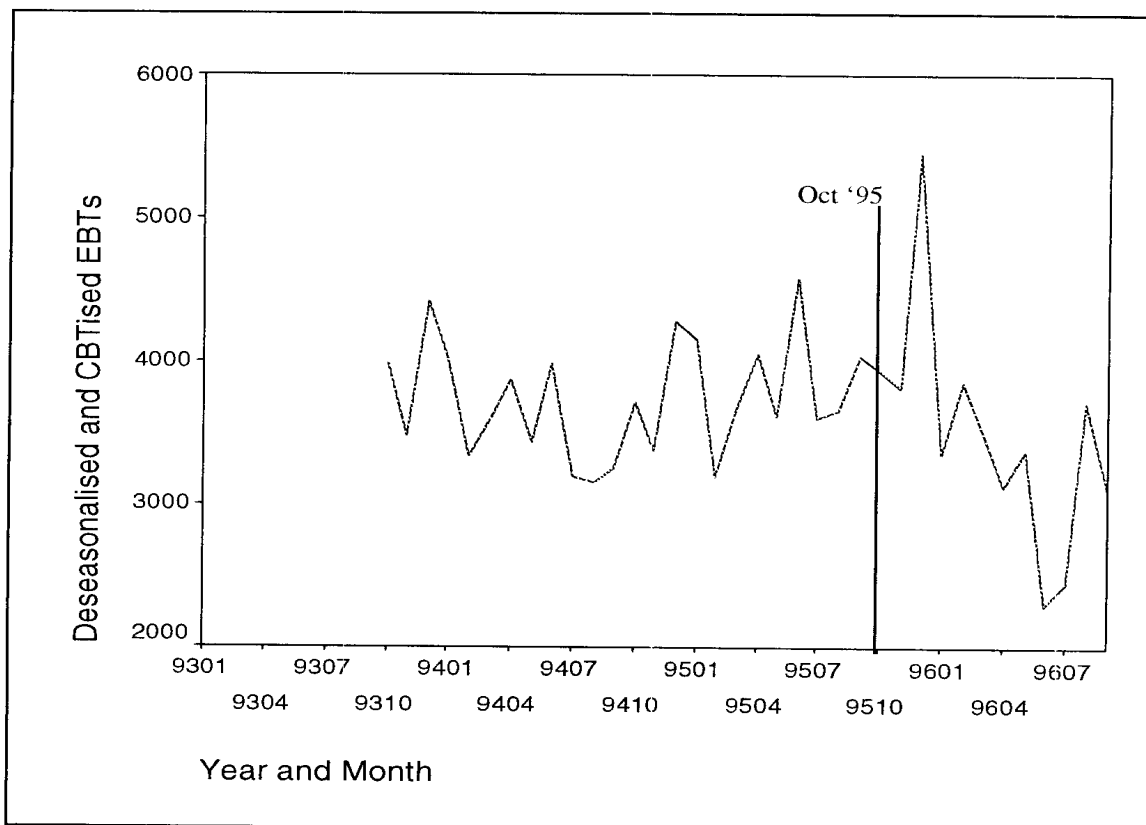
Figure 9. De-seasonalised EBTs also taking account of the variation in CBTs

Summary

This paper has attempted to explain why different experts get different results when analysing the same set of data. The paper tries to explain quite complex issues of data analysis with the help of diagrams so the explanation is accessible to the public who are not fluent in the specialised mathematical language usually employed by analysts. In its simplest form, the work that suggests that the drink-drive advertising campaign has been effective, has compared the period before the campaign began with the period after, and noticed an improvement in behavior.

In its simplest form, the work that suggests that the campaign has been ineffective, has looked at the data on a monthly basis and noticed when the monthly variation is removed, and account is taken of changes in CBTs, then there is nothing left to explain.

Each side of the argument has criticised the detail of the work of the other side in a technical way, but both sides have missed the point. The point is that it is impossible to know who is right, or if either is. The sort of data they are working with is not tractable to statistical analysis.

We have not given all the- alternative explanations for the results obtained so far, for instance it is easy to show that with the small addition to the Macpherson analysis, that allows for the idea that the period before October 1995 when the campaign was launched, is fundamentally different from the period after October 1995, then we can show that the campaign was very effective. The point about this is that just because an explanation make sense, does not mean

that it is correct. Another analyst with enough ingenuity can come up with a convincing alternative explanation that may also be supported by the data, again with enough ingenuity.

None of the critics has come up with the fatal flaw in the Macpherson paper which overshadows all the other technical criticisms. In the addendum to this paper this flaw is exposed, and a suggestion is made of how to overcome this flaw and also other problems.

Addendum and suggestion

In the body of this paper we showed that by thinking of a new way of looking at the data, we can come to a new conclusion. Of course, if the data does not fit with our new explanation, then we abandon it, and think of another. This process is called by analysts, modeling. What it has not done so far is to point to a more fundamental problem of analysing non-experimental data. Data are non-experimental if the analyst does not control the conditions under which the data are generated. The CBT and Adstock and other data are not set by the analyst, but arise from the desire of the government and the police to do something about the appalling road toll.

The rest of this section outlines the problem of analysing non-experimental data, by describing a hypothetical situation.

The police and the LTSA don't haphazardly schedule advertisements and breath tests. They advertise more heavily during the dangerous months, and make their presence on the roads more obvious when they think people will behave badly, in the hope of improving things.

Suppose the actual response to advertising is that a 1 % increase in advertising would lead to a 1 % decrease in EBTs. To make it more realistic, if advertising exposure is doubled then half of the people who would have drunk too much and then gone on to drive, would have stopped before they reached the limit, or have got a taxi home, or anyway would have not been driving over the limit. We don't know this number of course, but it is the number we are trying to get at when we do our analysis.

Suppose also that the police, for every 1% increase in anticipated EBTs, increase their advertising by I%. That is, when the police expect people to behave badly they increase their advertising proportionately.

Along comes an analyst who wants to use the data generated by this situation to measure the advertising response; that is, to estimate the effect of advertising on drink driving behaviour. What could the data look like? I will simplify the units in this hypothetical example so it is easy to follow.

Under this scenario, if the EBTs were 1000 in November and the police expected them to increase by 10% to 1100 in December because of the festive season, they would increase advertising and probably mount a blitz to reduce the bad behaviour. According to this hypothesised police response to an expected increase in drunk driving, they would increase their advertising by 10% in December. Just to continue the example, suppose they were

showing 1000 Adstocks of advertisements in November, then they would increase this to 1100, in order to combat the anticipated worsening situation.

The result of this increase in advertising, according to the advertising response model above would be to reduce the EBTs back down to 1000. The naive analyst plotting the data, would only notice a 10% increase in advertising from one month to the next, but no resulting decrease in EBTs, and conclude that the advertising was ineffective. Table 2 shows data that would be generated for: November and December.

Table 2. Two hypothetical situations when the police do not, and do, respond with extra advertising to an anticipated worsening of drink driving behaviour.

	This is what would happen with no increase in advertising		This is what would happen with a 10% increase in advertising	
	November	December	November	December
Advertising	1000	1000	1000	1100
EBTs	1000	1100	1000	1000

The response to advertising is hidden because we do not know what would have happened in the absence of advertising.

There is only one way to overcome this problem and that is to experiment with the advertisements; for instance show them in some reception areas and not in others, and analyse the results.

The author realises that there are technical issues that need to be carefully considered, but are confident they can be overcome. The author is not, however, confident that the ethical issues are ones that the public would feel confident with. He realises that where public safety is at stake, best practice should be followed. It turns out that the experts do not know what is the best practice.

References

Bliss T; Guria J; Vulcan P & Cameron M (1998). The New Zealand road safety advertising/policing package and its evaluation. *Proceedings: Road Safety Research, Policing and Education Conference* (pp. 1-4). -Wellington, New Zealand: Land Transport Safety Authority.

Cameron M & Vulcan P (1998). *Evaluation review of the Supplementary Road Safety Package and its outcomes during the first two years.* New Zealand: LTSA

Falconer W (1996, 31 January). Road safety: the challenge. *The Evening Post*, p.5.

Gregg S (1996, July 7). Victorian shock tactics drive home message. *The Sunday Star Times*, p.C3.

Macpherson T & Lewis T (1998). *New Zealand drink driving statistics: The effectiveness of road safety television advertising.* Marketing Bulletin, 9, 40-51

Tay R (1999). Effectiveness of the anti-drink-driving advertising campaign in New Zealand. *Road & Transport Research*, 8 (4), 3-15

White M (2000). *Re-investigation of data on the effectiveness of the first two years of the New Zealand supplementary road safety package.* (Chapter in Report Number 5/2000 White M; Walker J; Glonek G & Burns N (eds.). Adelaide, South Australia: Transport SA.

Tony Lewis is an Associate Professor in the Department of Marketing, Massey University.