Before-and-after studies using crash data and iRAP protocols
About iRAP

The International Road Assessment Programme (iRAP) is a charity dedicated to saving lives. Our vision is a world free of high risk roads.

iRAP works in partnership with government and non-government organisations to:

- inspect high-risk roads and develop Star Ratings and Safer Roads Investment Plans
- provide training, technology and support that will build and sustain national, regional and local capability
- track road safety performance so that funding agencies can assess the benefits of their investments.

Road Assessment Programmes (RAP) are now active in more than 50 countries throughout Europe, Asia Pacific, North, Central and South America and Africa.

iRAP is financially supported by the FIA Foundation for the Automobile and Society and Road Safety Fund. Projects receive support from the Global Road Safety Facility, automobile associations, regional development banks and donors.

National governments, automobile clubs and associations, charities, the motor industry and institutions such as the European Commission also support RAPs in the developed world and encourage the transfer of research and technology to iRAP. In addition, many individuals donate their time and expertise to support iRAP.

For more information

For general enquiries on this working paper, contact:

Dr Steve Lawson, Technical Director
International Road Assessment Programme
Worting House, Basingstoke
Hampshire, UK, RG23 8PX
Telephone: +44 (0) 1256 345598

To find out more about iRAP, visit www.irap.org. You can also subscribe to ‘WrapUp’, the iRAP e-newsletter, by sending a message to icanhelp@irap.org.

© International Road Assessment Programme (iRAP). iRAP technology including protocols, processes and brands may not be altered or used in any way without the express written agreement of iRAP.

iRAP registered office: 60 Trafalgar Square, London WC2N 5DS. Charity number: 1140357.

RAP504.11, May 2011
Contents

1 Summary................................................................................................................................. 4

2 Evaluation when implementing road safety countermeasures............................................. 4

3 Before-and-after evaluations ............................................................................................ 5

3.1 Before-and-after evaluations using crash data ............................................................... 5

3.1.1 Evaluations with treatment and control groups........................................................... 5

3.1.2 Regression-to-mean....................................................................................................... 6

3.1.3 Evaluations without a control group............................................................................ 7

3.1.4 Sample size.................................................................................................................. 7

3.1.5 Type of crash data........................................................................................................ 7

3.2 Before-and-after evaluations using intermediate safety performance indicators .......... 8

3.3 Examples of before-and-after evaluations by iRAP......................................................... 8

3.4 Statistical and economic analysis and reporting.............................................................. 11

4 Acknowledgements............................................................................................................ 12

5 References and further information.................................................................................... 12

Appendix: before-and-after study worked example................................................................ 14
1 Summary

Evaluation of any work is important because it demonstrates how effective that activity has been.

There are many ways to evaluate improvements to road infrastructure to check that they are reducing risk. It is possible to monitor safety at particular locations or over networks as a whole.

Over a long period of time, it is possible to check if the number of crashes has reduced at locations where infrastructure has been changed. A “before-and-after” methodology using crash data is described here. iRAP’s Risk Rate Mapping and Performance Tracking can also show before-and-after changes in risk and these techniques are illustrated here.

Where good crash data do not exist, it is necessary to measure risk indirectly, often by monitoring intermediate measures that are known to increase or decrease the risk of injury. The iRAP Star Rating is one such measure. Examples are shown of improvements in the safety of a road measured by the Star Rating.

2 Evaluation when implementing road safety countermeasures

At the core of iRAP’s work is the implementation of road engineering countermeasures at locations where severe injury risk is likely to be high and where the iRAP model demonstrates that deaths and serious injuries will prevented (and there will be economic benefits in investment).

Evaluation of all investment is important but is often forgotten or not done well. It is necessary to ensure that any change to the road system has the desired effect and to know the effectiveness of the engineering countermeasures used. More generally, there is a need to know that countermeasures are effective so as to maximise benefits from limited budgets and assess whether targets have been met. Evaluation also shows how a process can be improved, supports the business case for funding, fills gaps in knowledge – especially in low- and middle-income countries – and ensures that, at worst, no harm is done.

The process and the outcome can be evaluated. Process evaluations might focus on:

- how the countermeasure was delivered
- whether everything went to plan
- whether the methods were effective
- if scarce resources (such as staff time and budgeted funds) were used well
- how the process might be improved.

This note is mainly about assessing outcomes and looking at how effective the countermeasure was in reducing trauma and/or risk. This can then be used to check that the benefits of the countermeasure outweigh the costs.
3 Before-and-after evaluations

Several types of analysis can be used to evaluate the outcomes of countermeasures. Typically they involve a “before-and-after” study to assess either the number of crashes or some other measure of risk before and after the countermeasures have been implemented.¹

3.1 Before-and-after evaluations using crash data

Before-and-after evaluations can be at a single site or a group of sites, depending upon the extent of the engineering works.

3.1.1 Evaluations with treatment and control groups

Often before-and-after evaluations involve the comparison of a “treatment group” and a “control group”:

- Locations in the treatment group have countermeasures implemented and, numbers of crashes are measured before and after this occurs
- Locations in the control group do not have countermeasures implemented, and, numbers of crashes are measured for the same before-and-after periods as for the treatment group.

It is common that crash data for three or more years are used in both the before-and-after periods. Five years may give a more consistent picture of the long-term risk at locations but it is often important to assess before that length of time has passed. Data on before-and-after studies have been assessed and reported on in a number of ways – see, in particular, Elvik and Vaa (2009), UK MoRSE and the iRAP toolkit http://toolkit.irap.org/

¹ Randomised controlled trials are the “gold standard” in evaluation work. They are often used, for example, in drugs trials but are almost never used in infrastructure assessments, generally because of the complexity of doing so. In these studies it is the practice to randomly allocate individuals or groups to receive treatment. In this way it is possible to remove bias by balancing out all known and unknown variables between groups – the only remaining differences will be due to treatment. There is also a moral issue to this, in that, where it is known that treatment will reduce risk, it may be deemed immoral to choose not to do anything. Similarly, general statutory requirements to identify and treat known hazardous locations may sometimes mitigate against using control sites.
The treatment and control groups need to be closely matched so that their characteristics are similar to each other (see Appendix 1). The use of a control group that is well-matched with the treatment group means that general background trends, such as changes in traffic flow, are automatically “controlled” for (or taken account of) in the study, since these trends are assumed to be the same in both groups.

The effect of the countermeasure is thus evaluated as follows:

\[
\text{Countermeasure effect} = (\text{change in treatment group}) - (\text{change in control group})
\]

Where there is not close-matching of treatment and control groups such as that shown in Appendix 1, it is common for the number of crashes in the control group to be many times larger than the treatment group so that overall underlying trends such as seasonal variation or a general year-on-year reduction in the number of crashes can be accounted for.

### 3.1.2 Regression-to-mean

Crash numbers fluctuate from year to year such that, at a location after a high number of crashes in one year, the location might revert back to its “normal” number of crashes in the following year, regardless of whether a countermeasure has been implemented or not. This effect is often called “regression-to-mean” (RTM). It can lead to over-statement of the benefit of countermeasure and, dependent upon countermeasure type, it is not unusual for around 20-50% of the apparent reduction to come from the regression-to-mean effect. Statistical techniques that draw upon Bayesian statistics can be used to counter this effect. In recent work, Allsop (2010) explains this phenomenon and provides examples of evaluation of the effect of traffic cameras. Table 1 is from that study. The “Trend only” values show the percentage reduction in observed crash and casualty rates when only the underlying trend of generally decreasing crashes and seasonal variation is allowed for. “Trend and RTM” shows the reduction when both this trend and regression-to-mean is allowed for. The saving in road trauma that can be claimed is substantially less when “Trend and RTM” is allowed for.
Table 1  
Percentage reduction to numbers recorded from numbers expected at rates per year prevailing in the baseline period, after allowance for “Trend and RTM” (After Allsop, 2010)

<table>
<thead>
<tr>
<th>Type of camera site and effect allowed for</th>
<th>Percentage reduction between before-and-after periods (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Personal injury collisions</td>
</tr>
<tr>
<td>Fixed urban</td>
<td></td>
</tr>
<tr>
<td>Trend only</td>
<td>18.8</td>
</tr>
<tr>
<td>Trend and RTM</td>
<td>16.6</td>
</tr>
<tr>
<td>Mobile urban</td>
<td></td>
</tr>
<tr>
<td>Trend only</td>
<td>26.0</td>
</tr>
<tr>
<td>Trend and RTM</td>
<td>19.4</td>
</tr>
</tbody>
</table>

3.1.3 Evaluations without a control group

Before-and-after evaluations without a control group are often used in road safety evaluation. They measure the outcome of interest (such as a reduction in severe crashes) both before and after countermeasure and are inexpensive. However, they provide only weak evidence. They cannot attribute the change only to the countermeasure, as they often do not “control” for external (or confounding) factors, such as changes in traffic flows or broader road safety policies.

3.1.4 Sample size

An important part of any evaluation is the determination of an appropriate sample size. If the sample of data is too small, it will not be possible to detect a “statistically significant” difference (meaning that the difference is unlikely to have occurred purely by chance – see 3.4). If the sample is too large, it can increase the cost of the study, and might produce a significant result even if this is very small and not very meaningful within the context of the overall exercise. The appropriate sample size depends on likely size of the effect of the countermeasure, variability of a measure and the frequency with which the event occurs. Sample size calculators and more detail on any of these issues can be found online by entering those words in any search engine.

3.1.5 Type of crash data

If crash data are available, it can be disaggregated by fatal and injury crashes or by type (to see, for example, if a road surface treatment is having an effect on skidding crashes) or those related to speed, expressed as a rate or assessed for their severity (for example, the percentage of fatal crashes). Reliable crash data systems
are required. Under-reporting of crashes and even of fatal injuries is often a problem which should be taken account of.

### 3.2 Before-and-after evaluations using intermediate safety performance indicators

There are numerous situations where crash data is not available for use in before-and-after studies. This might occur because no reliable crash data system is in place (which is common in low- and middle-income countries) or where an evaluation is required before adequate crash data has accumulated after the road safety countermeasure has been implemented.

In these situations, “intermediate outcomes” or “safety performance indicators” can be used. The OECD report *Towards Zero* points out that intermediate outcomes or safety performance indicators provide a means of undertaking evaluations in the absence of good crash data (OECD/ITF, 2008). Measures such as helmet and seat belt wearing rates have been used effectively in assessing road safety behaviour, as have speed measurements and conflict studies, and iRAP provides a set of such safety performance indicators for road infrastructure.

iRAP involves the assessment of numerous road design attributes which are known to influence the likelihood and severity of crashes for vehicle occupants, motorcyclists, pedestrians and bicyclists. By recording a change in one of the attributes, it is possible to make an assessment of the likely impact on risk for that section of road. The iRAP approach is discussed in more detail in *Star Rating Roads for Safety: The iRAP Methodology*, which is available at: [www.irap.org/library.aspx](http://www.irap.org/library.aspx).

### 3.3 Examples of before-and-after evaluations by iRAP

Before-and-after evaluations are used within iRAP at several levels with its Star Rating protocol (see Figures 2 to 4) and in its Risk Rate Mapping and associated Performance Tracking (Figures 5 to 7). These protocols may be used to evaluate on the basis of an individual road section or for a whole network of roads.

In Figures 2 to 4, estimates of risk reduction before and after implementation of countermeasures are shown by changes in iRAP’s Star Rating. Justification of the safety benefit is based both upon evidence that the measures implemented have been demonstrated in other research to reduce deaths and serious injury, and by iRAP’s own research demonstrating a reduction in crashes with an increase in Star Rating (iRAP, 2011 (in preparation)). For road networks, the iRAP Star Rating process can be used to provide an estimate of the likely reduction in deaths and serious injuries based upon implementation of a “Safer Roads Investment Plan” over that network.
**Figure 2**  At this location in Johor, Malaysia, the installation of a pedestrian overpass reduced risk for pedestrians, increasing the Star Rating from 2-stars to 4-stars

**Figure 3** At this location in Pahang in Malaysia, improvements such as pavement resurfacing, roadside safety barrier, profiled edge markings and paved shoulder widening resulted in the run-off road risk for vehicle occupants reducing and the Star Rating increasing from 1-star to 2-stars

**Figure 4** At this location in Perak, Malaysia, shoulder sealing and improved delineation at the curve (road studs and hazard marker posts) reduced the run-off risk for vehicle occupants and resulted in improved rating from 1-star to 2-stars
Evaluation using iRAP’s Risk Mapping and Performance Tracking protocols applies a form of interrupted time series analysis where the crash rates on a population of roads is monitored. Evaluations may be made on the basis of:

- individual roads (see Figure 5)
- by a road authority interested in learning more about the efficacy of particular countermeasures (see Figure 6)
- at a national level by a government checking on its overall road safety engineering strategy (see Figure 7).

**Figure 5** Individual road sections showing greatest crash rate reductions are identified

**Figure 6** Detail of the evaluation from the two road sections that showed the greatest reduction
To produce these evaluations, data are initially analysed to identify individual road sections which have shown a reduction in the number of crashes over time and those where there has been little or no change. Figure 5 shows examples of typical road sections. Data for individual years are then checked to assess consistency of the patterns. Finally, highway authorities are asked for information on remedial, enforcement or education measures that have been implemented that might explain the reduction in crashes. Figure 6 shows how this evidence is used in presentation of the information.

3.4 Statistical and economic analysis and reporting

Most basic statistical textbooks will provide advice on the standard statistical tests that are applied in road safety. The main testing that most practitioners will encounter is in:

- the comparison of crash frequencies where a Chi-squared test may be used, or paired t-test if the distribution of crashes is assessed as coming from a Normal Distribution. (The Fisher Exact Test can be used instead of a Chi-squared test when any value in the cells of a 2-by-2 comparison matrix fall below 10).
- the comparison of crash rates using a paired t-test
- the comparison of proportions using a z-test.

In statistical analysis of crash reductions, the “95% confidence level” is typically used, although in some circumstances it is acceptable to use a 90% level (meaning that there is a 1 in 20 or a 1 in 10 chance respectively of the outcome occurring purely by chance).
Cost-benefit analysis is based upon estimating a value of life and injury and comparing it with the cost of intervention (the installation of countermeasures). These techniques are useful for providing a context for presenting a case for further funding and for allocating existing funding.

The basic report generally provided in evaluation will follow many of the examples shown here. It will describe the countermeasures being used, the types of locations and the populations being treated (for example, footbridges being used to reduce pedestrian risk). If a number of locations are being treated with a measure, then there will be details of the number of crashes in the before-and-after periods in both the treatment and control groups. The intervention effect will be reported in terms of change in crashes or a measure of risk or behaviour, often disaggregated by crash type and severity.

4 Acknowledgements

iRAP is grateful to Blair Turner, Principal Research Scientist, ARRB Group, for the presentation on which much of this summary is based, content that forms part of a two-hour workshop used by ARRB in low- and middle-income countries. Examples in Figures 2 to 4 are based on information supplied by Hamzah Bin Hashim, Assistant Director, Road Rehabilitation Unit, Road Facility Maintenance Branch, Public Works Department of Malaysia (JKR). Figures 5 to 7 were provided by the Road Safety Foundation. Additional advice was provided by TMS Consultancy (United Kingdom) and Peter Cairney (ARRB Group).

5 References and further information


The iRAP toolkit http://toolkit.irap.org/


UK Department for Transport Good Practice Guide
http://webarchive.nationalarchives.gov.uk/+http://www.dft.gov.uk/pgr/roadsafety/laguidance/roadsafetygoodpracticeguide with section on evaluation:

UK-MoRSE http://www.uk-morse.com/
6 Appendix: before-and-after study worked example

This example was provided by ARRB Group. Tables A1 and A2 show data for seven locations (sites T1-T7) where profile edge line was installed and the performance of this countermeasure was compared with control locations (C1-C7), usually on the same highway and adjacent to the treatment sites. Sites were well matched in terms of length and traffic flow, as can be seen from Table A1. They were also well matched in terms of geometry and adjoining land use.

![Profile edge line](http://toolkit.irap.org/default.asp?page=treatment&id=30)

The number of crashes of the type of interest (run-off-the-road and loss-of-control crashes) which occurred in the “before” and “after” periods are shown in Table A2, together with the results of statistical tests on the results of each site. In the case of the sites, this gives an indication of the occurrence of crashes in relation to the start and finish of the profile edge lines.

Table A2 shows that substantial crash reductions were experienced at both the treatment and control sites. The reduction at the control sites is likely to have been the result of ongoing improvements to the road system and the cumulative results of other road safety measures. It was assumed that any greater benefits at the treatmentcountermeasure sites would be attributed to profile edge lines.
### Table A1  
**Treatment and control sites – length, traffic flow and before-and-after periods – installation of profile edge lines**

<table>
<thead>
<tr>
<th>Treatment sites</th>
<th>Control sites</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highway</td>
<td>Length</td>
<td>AADT before</td>
</tr>
<tr>
<td>T1</td>
<td>9.7</td>
<td>3,000</td>
</tr>
<tr>
<td>T2</td>
<td>13.9</td>
<td>1,900</td>
</tr>
<tr>
<td>T3</td>
<td>5.8</td>
<td>21,000</td>
</tr>
<tr>
<td>T4</td>
<td>4.3</td>
<td>4,500</td>
</tr>
<tr>
<td>T5</td>
<td>2.1</td>
<td>4,500</td>
</tr>
<tr>
<td>T6</td>
<td>7.7</td>
<td>5,000</td>
</tr>
<tr>
<td>T7</td>
<td>1.3</td>
<td>9,500</td>
</tr>
</tbody>
</table>

### Table A2  
**Run off the road and Loss of Control Crashes at Treatment and Control Sites before-and-after installation of profile edge lines**

<table>
<thead>
<tr>
<th>Highway</th>
<th>Crashes before</th>
<th>Crashes after</th>
<th>Highway</th>
<th>Crashes before</th>
<th>Crashes after</th>
<th>Fisher Exact Test (or see footnotes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>21</td>
<td>2</td>
<td>C1</td>
<td>13</td>
<td>1</td>
<td>p = 0.46</td>
</tr>
<tr>
<td>T2</td>
<td>14</td>
<td>3</td>
<td>C2</td>
<td>3</td>
<td>2</td>
<td>p = 0.26</td>
</tr>
<tr>
<td>T3</td>
<td>20</td>
<td>2</td>
<td>C3</td>
<td>5</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>T4</td>
<td>12</td>
<td>2</td>
<td>C4</td>
<td>6</td>
<td>5</td>
<td>p = 0.09</td>
</tr>
<tr>
<td>T5</td>
<td>8</td>
<td>3</td>
<td>C5</td>
<td>7</td>
<td>3</td>
<td>p = 0.36</td>
</tr>
<tr>
<td>T6</td>
<td>23</td>
<td>12</td>
<td>C6</td>
<td>45</td>
<td>21</td>
<td>p &gt; 0.10</td>
</tr>
<tr>
<td>T7</td>
<td>10</td>
<td>3</td>
<td>C7</td>
<td>5</td>
<td>2</td>
<td>p = 0.39</td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td>27</td>
<td>--</td>
<td>84</td>
<td>34</td>
<td>p = 0.051</td>
</tr>
</tbody>
</table>

1. Fisher Exact Test not applicable due to 0 crashes in control section during “after” period.
2. Chi-squared test applied since numbers are sufficient. Chi-square value = 0.062, df = 1, p > 0.10
3. Chi-squared test applied since numbers are sufficient. Chi-square value = 2.673 df = 1, p=0.051
Statistical testing was carried out to test whether the crash reductions at the treatment sites were greater than the reductions at the control sites. The Fisher Exact Test was used to test for differences at each pair of sites. Like the Chi-squared test, this test is used where there are nominal level data but, unlike the Chi-squared test, it is suitable where cell frequencies are less than 5 (Siegel, 1956). The test is used here to determine whether the distribution of crashes between the “before” and “after” is the same at the treatment site is the same as at the control site.

The results of these tests show that in only one case was there a significant difference between the treatment and control site – at T4 where there was a very marked reduction and where there was almost no change at the control site, C4. At T3-C3 the test was not applied because there were no crashes in the after period at the control site and therefore no possibility of a significant difference in the hypothesised direction. The Chi-squared test was applied to the T6-C6 comparison because there were sufficient numbers in each cell of the comparison to allow this test rather than the Fisher Exact Test. The test, however, was not significant (Chi-square = 0.06 df = 1). ARRB report that the proper conclusion to draw from this part of the analysis is not that the treatments have been ineffective but that the crash numbers are too small at individual sites to allow statistical conclusions to be drawn.

The results from all treatment and control sites were then combined (using the column totals in Table A2 and a Chi-squared test was applied to these aggregated results. In this case a Chi-square value of 2.67 was found, with 1 degree of freedom. This is significant at the 0.1 level on a 1-tail test (the critical value of Chi-square is 1.64) and just falls short of being significant at the 0.05 level (critical value = 2.71). Thus, according to current statistical convention, there is not a significantly greater reduction in crashes at the treatment sites than at the control sites, although it is clear that a slightly less demanding statistical criterion would indicate that the treatment had benefits.