A Proposed Framework for Evaluation of Road Safety Strategy Outcomes

Stuart Newstead (Presenting Author) and Kathy Diamantopoulou
Monash University Accident Research Centre
Email: stuart.newstead@monash.edu

Abstract

Road safety strategies have been developed and implemented in jurisdictions across Australasia to focus efforts and resources with the goal of achieving set targets for road trauma reduction. Evaluations of specific programs implemented as part of road safety strategies are common. Rarely however is the success of the road safety strategy as a whole evaluated in a comprehensive and systematic manner. This paper details the development of a comprehensive framework for road safety strategy outcome evaluation. The evaluation framework proposed is based on the GOSPA framework of defining a road safety strategy formulation. The GOSPA framework defines a pyramid of increasing detail in defining the elements of a road safety strategy. The top of the pyramid defines the broad goals for which the strategy is aiming (Goals) whilst the next level down gives specific measurable targets (Objectives) against which the goals can be assessed. The Strategies area of the framework typically defines the target areas on which the road safety strategy will focus to achieve its goals along with local objectives within each target area that will jointly contribute to achieving the global objectives. Finally the Programs and Actions areas contain the specific details on the type of activities to be carried out in each target area and the amount of effort that will be applied to each activity.

The evaluation framework developed mirrors the pyramid structure defined under the GOSPA framework. It is designed to assess the progress of the strategy against the pre-determined goals and objectives at various levels of detail through a multi-tiered modelling approach. Each proposed tier of evaluation focuses on a specific level of disaggregation of the strategy elements corresponding to particular levels of the GOSPA framework pyramid. Examples of application of the framework are demonstrated on the 1993-2003 Queensland Road Safety Strategy. Its potential for general application to other road safety strategies is discussed along with the potential use of the framework for forward planning of evaluation activity including resources allocation and data requirements.

Keywords:

evaluation, road safety strategy, crash, injury, statistical analysis

Background and Aims

Road safety strategies play a major role in reducing death rates and injury rates resulting from road trauma. Such strategies consist of a range of individual road safety programs, each contributing a varying level of success towards reducing crash rate/frequency and crash severity.

To gain an understanding of how successful a road safety strategy has been is not as simple as looking at the numbers of deaths and injuries to see whether they have increased or decreased over a certain period. Any number of explanatory variables may have had an impact on the outcomes of a road safety strategy. Changes in the economy, increased population and increased motorisation all have the capacity to affect the level of success achieved in reaching road safety targets [1]. Evaluation plays a vital role in assessing which aspects of a road safety strategy work and which aspects need to be reviewed. Such a tool can help reallocate funds and resources towards programs which will have the greatest impact on reducing deaths and injuries.

The more commonly reported method of road safety evaluation to date is individual program evaluation. Examples of this throughout Australia include programs related to speed cameras, random breath testing, moving mode radar, accident black-spot treatments, 50 km/h speed limits and Random Road Watch. With information on the effectiveness of each individual program, estimates can be made of the number of lives and injuries saved/prevented as a result of each program. However, there may be difficulty in measuring the interaction between individual programs and their respective contribution to the overall strategy goals. For example, a reduction in speed-related crashes could be influenced by speed cameras, 50 km/h speed limits and mass media advertising, yet the extent to which each of these programs contributed to the outcome is often much less clear.

Road safety strategies implemented across Australian jurisdictions have had varying degrees of evaluation undertaken to assess what effect the various programs and the strategy as a whole had on road safety. Whilst there are examples of individual program evaluations, few were found to evaluate the entire strategy with consideration not only to each of the programs, but to the many factors influencing road safety outcomes. Much of the work is unpublished and often takes the form of reports to Government departments or related agencies.

The aim of this study was to develop a general evaluation framework to monitor and report on the outcomes of a typical road safety strategy over the period of its operation. Requirements of the framework were that it should assess the effectiveness of the strategy both at a global level and within specific target areas and that it would take into account the many factors influencing road safety outcomes.

A Theoretical Framework in Which to Frame a Road Safety Strategy

To develop an effective road safety evaluation framework, it is first necessary to define a theoretical framework in which the strategy can be expressed in order to define the levels and activities for evaluation. A review of the literature identified two potential theoretical frameworks under which a road safety strategy could be expressed. They were the GOSPA framework [2] and the PIARC framework [3].

Assessment of the most appropriate of these models for use in this study was based on the consideration that comprehensive strategy evaluation needs to be undertaken at a micro level (to assess the effect of any individual road safety initiative) as well as at a macro level (to assess the general effect of a group of initiatives) to ensure that the effect of all road safety initiatives within the strategy is measured both collectively and individually. There is also a need for global-level evaluation in which the individual effects of many (if not all) of the initiatives are modelled collectively, in order to assess the effects of those initiatives which cannot be subjected to micro-level evaluation for a variety of reasons. Based on a comparison of the two frameworks against these criteria, it was decided that the most appropriate theoretical framework on which to develop a comprehensive evaluation strategy was GOSPA. Furthermore, this framework reflects that most road safety strategies in Australia are structured in terms of an overall goal, objectives, strategies, programs and actions

The structure of the GOSPA framework is as follows:

Goal	General (idealistic) statement of the Program's overall goal
Objectives	Specific (pragmatic) statements of the Program's measurable objectives to reach the Goal
Strategies	General (idealistic) strategies to achieve each Objective
Programs/Plans	Specific (pragmatic) programs/plans contributing to each strategy, with measurable activity levels and outputs
Actions	Actions undertaken in each program.

A typical road safety strategy and associated action plans can be expressed to fit the GOSPA framework as illustrated in Table 1 where typical examples of the strategy are given under each heading. The GOSPA model allows an evaluation structure to be defined. Programs are not considered to be implemented unless actions are taken, strategies are not achieved unless planned programs are implemented, objectives are not met unless the strategic directions of programs are correct, and goals are not achieved unless the targets of objectives are met. Through this strategic framework, the key pathways through which each action, program and strategy contribute to the overall goal becomes apparent. Measurable criteria at each level provide the basis for assessment that real change has occurred, and potentially developing linkages to the overall goal through modelling of the linkages.

Table 1: *Road Safety Strategy structured under the GOSPA Framework*

GOSPA Framework Component	*Examples*
Goal	• To minimise crash severity and reduce long-term consequences of injuries
Objectives	• To achieve a reduction in the number of fatalities to less than 5.6 deaths per 100,000 people by the year 2011, and to achieve a reversal in the increasing trend in hospitalisation casualties and the hospitalisation rate.
Strategies	• Safe attitudes and behaviours and optimal health outcomes in the event of a crash; safe roads, safe road environments and safe management of traffic. • To target broad key road user groups and behaviours, e.g.: Drink-drivers; speed-related crashes and speeding drivers; fatigued drivers; young adult inexperienced drivers; older road users; intoxicated pedestrians; unrestrained occupants; unlicensed drivers and riders.
Programs	• Police enforcement programs targeted at drink-driving and speeding; public education campaigns targeted at fatigue and speed.
Actions	• Increase hours of random breath test operations and speed-related enforcement operations; implement countermeasures to address drink-walking; trial the use of automatic number plate recognition technology targeted at unlicensed drivers.

The Proposed Evaluation Framework: A Multi -Tiered Evaluation Approach

A primary consideration in developing a comprehensive road safety strategy evaluation framework was that it should ideally provide information on the extent to which:

- each action contributes to the relevant program meeting its objectives,
- each program contributes to the relevant strategy meeting its objectives,
- each strategy contributes to the overall goal and its target and
- the overall objectives of the strategy have been met.

The GOSPA framework defines a pyramid of increasing detail in defining the elements of a road safety strategy. The evaluation framework that has been developed mirrors the pyramid structure defined under the GOSPA framework. It is designed to assess the progress of the strategy against the pre-determined goals and objectives at various levels of detail through a multi-tiered modelling approach. Each tier of modelling focuses on a specific level of disaggregation of the strategy elements corresponding to particular levels of the GOSPA framework pyramid. These are described broadly as follows:

- **The Global or First Tier Assessment Model**: This level of assessment aims to measure the ongoing performance of the road safety strategy in achieving the broad **G**oals and **O**bjectives set out for the strategy as a whole. Specifically it aims to measure whether the strategy has reduced overall road trauma levels, as defined by measures specified in the strategy Objectives, from that expected based on pre strategy implementation trends.
- **The Second Tier Assessment Model:** This level of assessment has aims similar to the First Tier models except here assessment is aimed at each individual target strata defined by the **S**trategies of the Queensland Road Safety Strategy and corresponding Program areas. Assessment at this level will aim to articulate in what particular areas the strategy is working and to what degree.
- **The Third Tier Assessment Model:** This level of assessment aims to relate the trends observed in each of the target strata defined in the Second Tier models to explicit measures of road safety program for major activities defined in the **P**rogram and **A**ction elements of the strategy.

Table 2 depicts how the GOSPA framework relates to the three-tiered models and fits in with the test-running of the framework on the earlier strategy.

Table 2: *Summary of the link between the GOSPA framework and the tiered evaluation framework*

GOSPA FRAMEWORK COMPONENTS		MODELLING TIER
Component	**Definition**	
Goal	Overall goal of strategy (i.e. to prevent road trauma through safe road use, safe roads and safe vehicles)	Global assessment model (top-tier model) to measure effect on road trauma of the Strategy overall
Objective	Objectives to reach goal (e.g. to achieve a reduction in the fatality rate to under 5.6 deaths per 100,000 people)	
Strategies	General strategies to achieve objectives given in the Queensland action plans and road safety strategy	Second-tier modeling of specific strata targeted by the strategies in the action plans (e.g. crashes occurring during high alcohol times of the week)
Programs	Specific programs relating to target group outcomes	Third-tier modeling of the individual program elements of the strategy
Actions	Actions undertaken in each program	

A fourth tier of the evaluation framework is also included which describes the imperative of undertaking specific targeted evaluations of major road safety programs implemented or enhanced under the road safety strategy. Such evaluations will give the most rigorous scientific assessment of the effectiveness of the programs evaluated and contribute to understanding of the mechanisms of effectiveness. Furthermore, results of the specific evaluations are critical in assisting the formulation of the Tier 3 models by identifying the most relevant measures of program activity that predict the outcomes being measures.

Detailed Description of the Framework Elements

The Global Assessment or Top Tier Model

The Global Assessment Model focuses on assessing the broad goals and objectives of the road safety strategy. It evaluates the overall effects of the strategy on the key outcomes defined (generally crashes or crash rates at particular severity levels). The Top-Tier model includes assessment of variations in major socio-economic factors across time and regions of the jurisdiction and their relationship with the outcome measures in order that these influences, which can potentially contaminate measurement of the effects of the strategy, will be removed. Format of the data for the Tier 1 evaluation models will generally be time series data of the key outcome measures both before and after implementation of the strategy. Correspondingly, statistical models to facilitate the Tier 1 evaluation will generally be based on time series methodology that allows for the inclusion of covariates into the model structure and provide simple, accurate forward forecasts of trends. Statistical state-space models are an example of such a modelling technique[4, 5].

The global assessment model can be formulated in two different ways depending on the time frame in which the evaluation is conducted formulated relative to the implementation of the strategy. In the case of formulating the evaluation model at the commencement of the strategy, a time series model is estimated that models the key outcome measure by month before the introduction of the road safety strategy as a function of other key non-road safety program related covariates. It is then used to forecast the levels of road trauma that would have been expected (together with confidence limits on the estimates) to have occurred after strategy implementation had no strategy been in place (based on the prior trends). Against the forecasts from the resulting model can be plotted the actual road trauma trends that occurred after strategy implementation. Plotting of the actual trends against that forecast in the absence of the strategy can be easily achieved by lay staff without statistical training as a means of monitoring overall strategy performance over time. Figure 1 depicts this graphically.

The time-series model that will investigate road trauma trends pre and post strategy against a forecast trend post strategy in the absence of the strategy are similar in philosophy to control-chart methodology used by a number of agencies in the past to monitor road safety strategy performance [6, 7]. The advantage of the proposed methodology for the global assessment model here is that it employs much more sophisticated and robust statistical methodology yet is still amenable to use by those without statistical training once established.

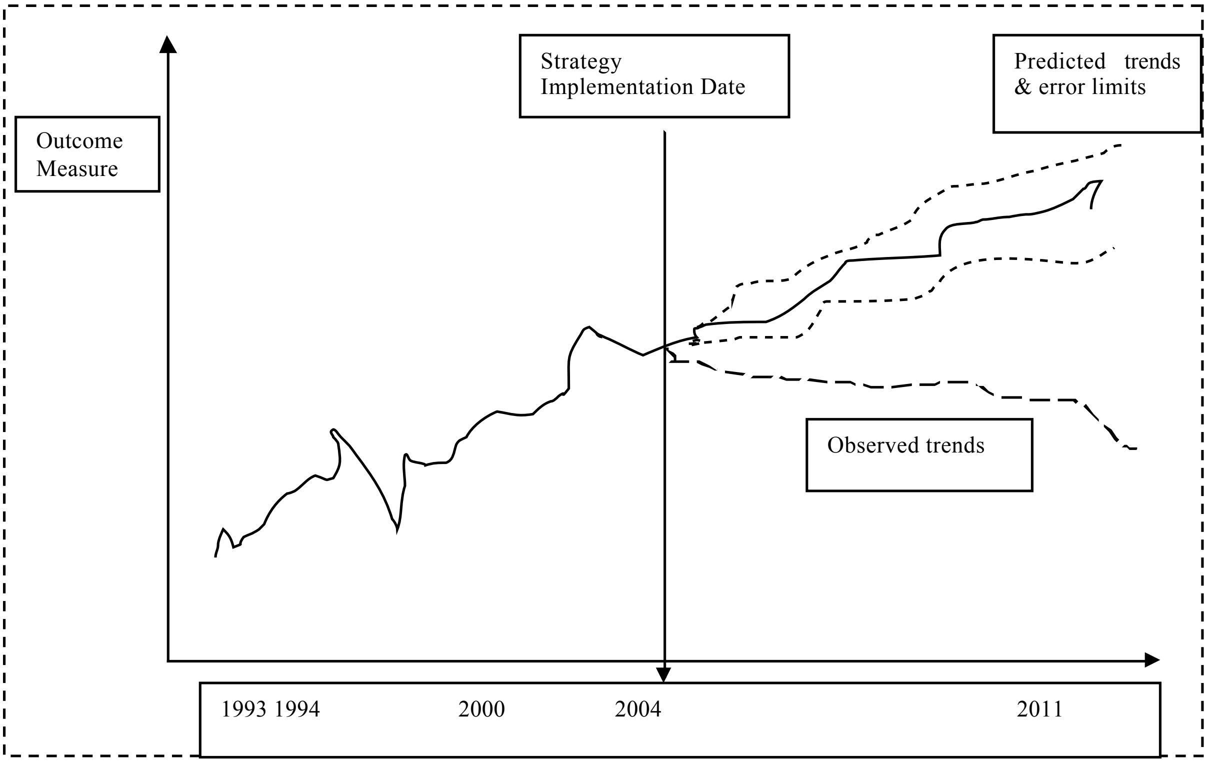

Figure 1: *Model of observed and predicted road trauma trends pre and post the introduction of the road safety strategy*

The global assessment model can also be used at various time periods after strategy implementation to formally evaluate the overall performance of the strategy to that point in time based on the key outcome measures specified in the Strategy's goals and outcomes. Here the pre and post implementation data to the time available are modelled with an intervention term being included at the time of strategy implementation. The intervention term parameter then represents the effect of the strategy on the outcome measure which can be tested formally for statistical significance. The intervention term can be modified accordingly to reflect increasing effects of the broad strategy over time which might be expected if components of the strategy are introduced in a staggered manner over time or take some time to become fully effective. Application of the global assessment model in this manner would require high level trained statistical expertise.

The Second Tier Model

The second tier model relates to the objectives defined by the strategies and programs of the road safety strategy and associated action plans. This evaluation level considers the effects on road trauma for specific strata defined by the road user groups or situations (e.g. pedestrians; young drivers; high alcohol crashes) at which the strategies and programs are targeted.

Examples of broad key road user groups and behaviours/situations that would typically be targeted in the second-tiered model include:

- Alcohol and drug-driving (modelled during high alcohol times of the week)
- Speed-related crashes and speeding drivers
- Fatigued drivers
- Young adult inexperienced drivers
- Older drivers

- Fatal and serious crashes in rural Queensland
- Pedestrians, including intoxicated pedestrians
- Unrestrained vehicle occupants
- Unlicensed drivers and riders
- Motorcycle riders

For each of these strata, a specific analysis model equivalent in structure to that defined by the Tier 1 model above would be estimated for each of the key outcomes being measured. Like the global assessment models, the second tier models can be formulated at time of implementation to forecast road trauma outcomes in each stratum of interest had the strategy implementation not taken place. Actual post implementation road trauma trends are then compared to those forecast to assess strategy effectiveness. Intervention models can also be estimated at time points after strategy implementation to formally assess the statistical significance of outcome changes related to the strategy for each strata defined above. Expertise required for each approach is the same as for the global assessment model.

The Third Tier Model

The third tier model relates measures of the individual program elements and actions of the road safety strategy and action plans (for example; the effects of speed camera operations) to the variation in outcome measure in order to assess the specific association between action and outcome. The Tier 3 modelling strategy is an extension of the Tier 2 model in that it will typically target the same strata defined in Tier 2. However, instead of modelling historic trends through general terms in the time series model, the model includes specific measures of road safety program effort under different activity areas as model covariates along with the non-strategy related covariates already included. In this way, the model makes estimates of the effects of individual initiatives (where there is sufficient data for the estimate to be reliable) by establishing the association between measurable road safety program effort and the key strategy outcome measures and relating the real variation in program effort to the reduction in road trauma observed. Results from the Tier 3 modelling process give specific estimates of the relative contributions of each of the major program elements in the road safety strategy to achieving the measured outcomes.

Poisson or negative binomial regression models are typical of the types of statistical techniques used in this third tier modelling approach. They are fitted to the outcome data series at some point after program implementation when sufficient post strategy and program element implementation experience has been accumulated to allow for successful modelling outcomes. The Tier 3 modelling process must be carried out by someone with high level statistical expertise.

There will be some types of initiatives whose impact on crashes cannot be assessed using the Tier 3 modelling approach due to relatively small target group and/or duration of operation or because program element input cannot be measured in a meaningful way. In some instances these program will be represented as local interventions. Where this is not possible, the aggregate effects of such programs can be assessed through comparing the third tier modelling outcomes with the aggregated effects accounted for in the global assessment model. Examples of types of road safety activities that can be included in the Tier 3 evaluation models include: hours of police enforcement by various modes such as speed cameras, number of random breath test conducted, media publicity levels achieved, proportion of the vehicle fleet with 5-star NCAP performance, number of accident black spots treated, etc. Table 3 illustrates further examples of program output measures and the Tier 2 defined analysis strata that they relate to.

Table 3: *Examples of programs with appropriate input variables to be used in explanatory models with key strata for categorising outcome measures*

Examples of programs	Input variables	Key strata
Implement best practice planning and design of cycling and pedestrian facilities for local roads.	Number of facilities with these features per time period	Pedestrian and cyclist injuries; local roads
Provide at least 170,000-190,000 police officer hours towards random breath testing over each year of the Action Plan.	Number of tests done per time period; number of hours of enforcement done per time period	High alcohol hours crashes; crashes specific to region and hours of enforcement specific to region
Advertising campaign to promote the dangers of driving tired, or without due care and attention.	Measures of publicity intensity (e.g. TARPS) per time period; advertising air time per time period	Region level crashes if exposure to advertising different per region
Undertake key enforcement activities such as the delivery of at least 43,800 hours of speed camera activity	Enforcement hours per time period	Crashes specific to region and hours of enforcement specific to region; low alcohol hours crashes
Introduce new penalties and sanctions for non-use of restraints.	Date when penalties were first publicised; hours of dedicated enforcement per time period	Crashes specific to region and hours of enforcement specific to region
Trial the re-introduction of L-plates on all vehicles driven or ridden by holders of learner licenses.	Date of introduction of measure	Young or learner driver crashes

Tier 4: Specific Evaluation

The final tier of evaluation in the framework is specific evaluation of major program components. The Tier 3 models described above measure the general association between measures of specific road safety program activities in a multivariate setting. However, for large complex road safety program elements, specific evaluation of major elements is generally needed for two reasons. First, only specific evaluation can establish the cause and effect relationship between road safety program element implementation and road trauma outcomes with a sufficient degree of scientific rigour. Second, specific evaluation is often needed to establish the measure of road safety program operation that is best related to the outcomes achieved which in turn is fed into the Tier 3 models as a key input. For these reasons, it is vital that specific evaluations of key road safety programs implemented as part of the broader road safety strategy continue to be carried out.

Demonstration of the Evaluation Framework

To demonstrate how the evaluation framework works in practice, a test-run of the framework was trialled on the previous 1993-2003 Queensland Road Safety Strategy. Sufficient crash data prior to 1993 was not available to support the modelling process so it was decided to trial the evaluation framework on assessing the outcomes of the revision to the strategy implemented from 1998. Evaluation models were fitted to the Queensland police reported crash data for the period April 1991-December 2004. As per the evaluation framework specification, explanatory factors were also included in the models to improve fit and predictive power as well as give insight into how changes in these factors might affect forecast trends. Factors included were the monthly unemployment rate and monthly fuel sales for Queensland.

The 1993-2003 Queensland Road Safety Strategy was launched in April 1993. It included an outline of the vision of road safety for 2003, and also outlined the objectives and principles to achieve that vision. The strategy contained more than 120 specific actions of which 85% had been implemented by 1998. The target outcome of the strategy was for actions to lead to a 30% reduction in fatal crashes relative to the trend by 2003. This target was met in 1998 with an estimated 8.07 fatalities per 100,000 population recorded. This was a 42%

reduction on the 1992 fatality rate of 13.7 deaths per 100,000 population. In addition there were improvements for most road user groups and crash categories. Because the target of a 30% reduction was met by 1998, a revised Strategy was launched in 1999. The main target of this revised Strategy was to achieve a 20% reduction in the annual fatality rate by 2003 compared with the 1998 figure.

First Tier Evaluation Models

Queensland crash data for the period April 1991-December 2004 was considered in the modelling process. Consistent with the stated outcomes of the strategy, fatal crash risk per 100,000 population was modelled, with fuel sales and unemployment rate included as covariates. Crash risk was modelled for the 'before' period April 1991 to December 1996 – using quarterly crash risk figures for fatal crashes. Crash risk trend was then forecast from 1997 to 2001 along with 68% confidence limits. Sixty-eight percent confidence limits were chosen to be consistent with control chart analysis techniques [8]. The models included covariates unemployment rate and fuel sales. State-space models were used including the covariates with the crash risks transformed into natural logarithms before being modelled. The modelled and forecast fatal crash risks are shown in Figure 2, together with what was actually observed in the period of the strategy 1997-2001.

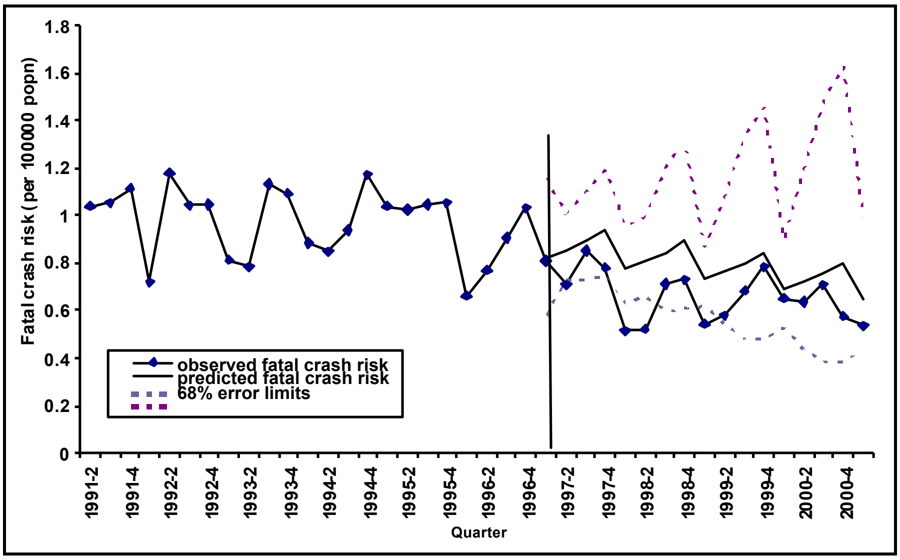

Figure 2: *Queensland Fatal crash risk per 100,000 population per quarter, April 1991-March 2001*

As can be seen in Figure 2, the observed fatal crash risk during the period of the revised strategy was somewhat lower than that predicted from the model indicating the success of the strategy. Simply using the observed crash trends show the revised strategy was successful in reducing fatal crash rates by the estimated 20%. The Tier 1 modelling however indicates that at least some of this success was driven by factors not related to the strategy with the real success of the strategy being somewhat smaller. Figure 2 also demonstrates the high inherent variability in fatal crash risks as indicated by the relatively wide confidence limits on the forecast.

Intervention Model

Post hoc assessment of the intervention effects of the revised Queensland strategy fatal crash risk was also carried out under Tier 1 of the evaluation framework. Here the model was fitted to the full data from April 1991 to December 2001 again including unemployment rate and fuel sales as covariates and also including annual intervention terms to measure the annual effectiveness of the strategy. Figure 3 shows the resulting model (Level + Reg + Intervention) and observed data whilst Table 4 gives the estimated net annual fatal crash reductions associated with the strategy from one year to the next together with their significance probability.

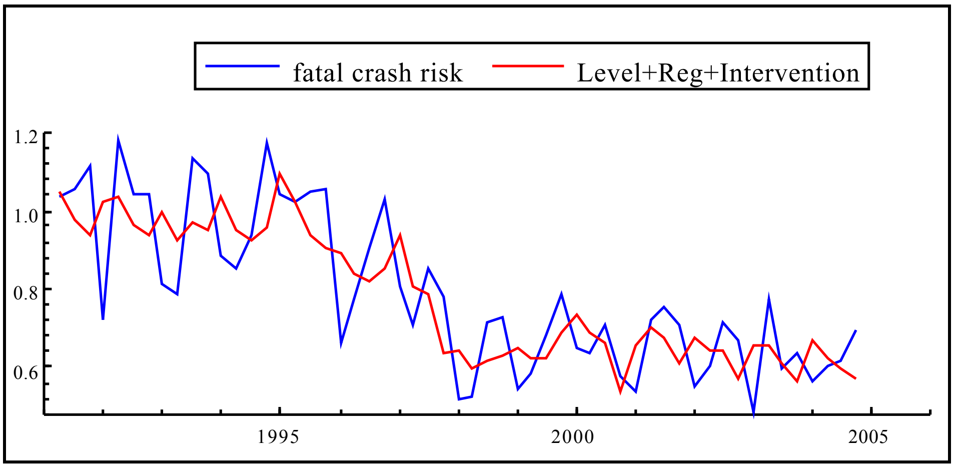

Figure 3: *Queensland fatal crash risk intervention model*

Table 4: *Estimated fatal crash risk reduction associated with the strategy*

Post Strategy Year	Crash Reduction From Previous year	Significance probability
1997	-5.21%	0.72
1998	23.61%	0.09
1999	12.37%	0.36
2000	12.53%	0.35
2001	5.97%	0.67
2002	-6.49%	0.94

The results in Table 4 indicate that there has been a period of sustained reduction from about 1998 to 2001 before levelling off again after that. Although the step intervention functions fitted to the model were not statistically significant they are indicative of a reduction in the fatal crash risk.

Second-Tier Evaluation Models

A number of key road user groups or behaviours targeted in the revised strategy including drink drivers, speed related crashes, rural road users, older road users and young drivers. An evaluation modelling approach was carried out for each target group in Tier 2 of the framework as per the techniques demonstrated above for Tier 1. For brevity, only an example of a Tier 2 model outcome is shown in Figure 4 which gives the analysis results for serious casualty crash risk in high alcohol times in regional Queensland. The analysis indicates the success of the strategy in lowering high alcohol hour crash rates indicated by the observed crashes being lower than the forecast and generally outside the control limits on the forecast.

In the later months of the prediction period for many of the Tier 2 models fitted, the confidence limits on the pre strategy trend projection typically become very wide meaning statistical comparison of the observed data with the forecast became problematic. This is particularly the case in the period more than 2 years after strategy implementation and suggests that comparisons with the forecast should be limited to 2 years of forecast data. In practice, this means that the forecast will need to be re-estimated periodically, for example every 2 years for the examples considered, against which to compare the ongoing success of the strategy. Accuracy of the initial forecast data will rely somewhat on the length of pre-strategy data available for analysis with longer time periods giving more accurate forecasts. The time period at which forecasts need to be re-estimated will depend on how much prior data the initial forecast is based. Similar comments on forecasting accuracy are also relevant to the global assessment level models.

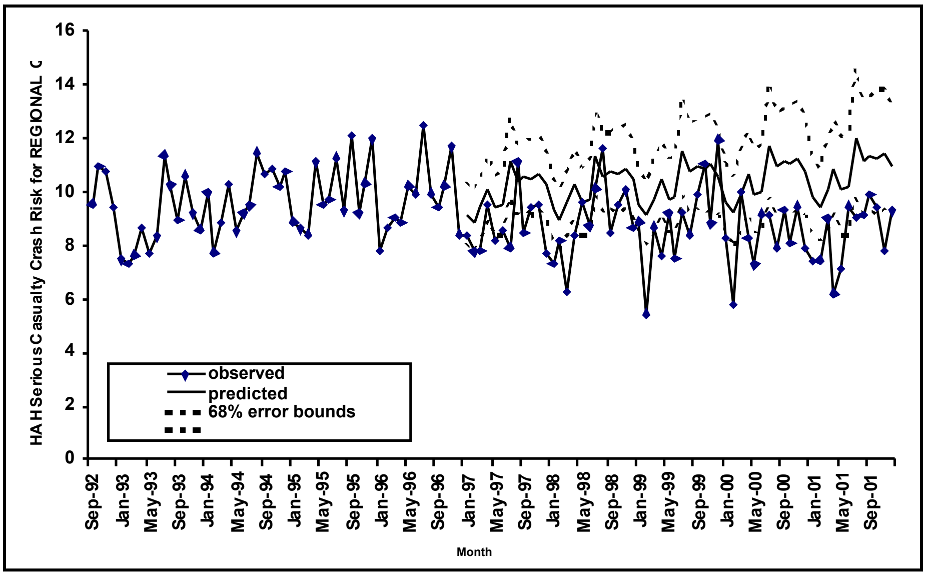

Figure 4: *Serious Casualty Crash Risk during High Alcohol Times of the Week in Regional Queensland*

Third Tier Evaluation Models

Application of the Tier 3 modelling approach to the previous road safety strategy period in Queensland has been demonstrated previously in evaluating the effects of the Road Safety Initiatives Package (RSIP) implemented as part of the revise Queensland road safety. Undertaking Tier 3 analysis such as in the RSIP example given is a complex process requiring high level statistical expertise to design and execute the analysis successfully. Careful detailed planning is required to successfully construct such models along with reasonable a-priori knowledge of the road safety programs considered in the analysis in terms of the most appropriate measures of the key operational elements leading to their success. It was beyond the scope of this paper to fully describe the Tier 3 model outcomes but a full description of the evaluation and the RSIP can be found in [9].

Fourth Tier Evaluation

The fourth tier of evaluation is specific evaluation of major program components. For the revised Queensland road safety strategy, specific evaluation was undertaken for a number of key programs. Again, it is beyond the scope of this paper to describe these fully. Furthermore, the techniques used for each specific evaluation will be highly dependent on the specific evaluation being carried out so it is not possible to give a general summary of methodology. Some specific examples of Fourth Tier evaluations related to the revised Queensland strategy are given in Table 5 along with the type of designs used, data required and estimated program effects.

Table 5: *Examples of specific program evaluations in Queensland*

Program Evaluation	Reference	Design	Data requirements	Estimated reduction
An evaluation of the 50km/h default speed limit in regional Queensland	[10, 11]	Time series analysis of intervention (50km/h areas) with control areas (60-70km/h)	Crash data; Speed survey data; separate analysis for young drivers, older drivers, pedestrians	13.5% for all crashes; 19.3% for fatal, serious and medical attention crashes combined
Evaluation of crash effects of the Queensland speed camera program	[12]	Time series analysis of intervention (areas within 6km of the camera site) with control areas (areas further than 6km)	Crash data inside and outside speed camera zones; number of speed camera zones, sites, site density and hours of operation by police region	32% reduction in fatal crashes, a 26% reduction in fatal to medically treated crashes combined and a 21% reduction in all reported casualty crashes

Evaluation of the Queensland Random Road Watch Program	[13]	Analysis of treatment and control site crash rates before and after intervention, stratified into rural/urban and region	Operational details of program; crashes by strata defined and by severity level	11% crash reduction in aggregate; 13% for serious injury crashes; 9% for damage only crashes

Discussion

This study has defined a framework for the global evaluation of road safety strategy outcome measures and has demonstrated the practical application of the framework on the 1996 revision of the 1993-2003 Queensland road safety strategy. The test run of the framework proved its efficacy and highlighted the benefits that can be gained from defining a systematic evaluation framework. It also highlighted a number of aspects that need to be treated with some care and some practical limitations of the approach.

Through the application of time series analysis techniques including covariates, the Tier 1 and 2 evaluation levels offer the opportunity to assess the net effects of a road safety strategy against other competing factors influencing road trauma outcomes. An advantage of these evaluation models is that the data on which they are estimated is generally readily available. Furthermore, inclusion of non-strategy related covariates in the models gives a measure of how these factors can affect road trauma outcomes and hence how these factors can work for or against achieving objectives set in the strategy.

Time series forecasting models used in Tier 1 and Tier 2 of the framework need to be formulated with care. To achieve the best results from these analysis as many key drivers of road trauma outcomes not related to the strategy being evaluated need to be included to ensure accurate forecasts in the pre-hoc evaluation models and unbiased measures of the strategy intervention effects in the post hoc application. Pre-hoc application of the Tier 1 and 2 framework models gives a ready means by which road safety agencies can track strategy performance in real time without high level statistical skills once the models are established. However, the test run of the framework shows that the accuracy of the forecasts diminishes rapidly the further forward the forecast is projected. In addition, the forecasts are based on static projections of the covariates included in the models meaning the forecast could be invalidated if there is a significant change in these covariates during the post strategy implementation period. The potential for change in the covariates increases with the length of the forecast. For these reasons, in practice the forecasts used for the pre-hoc application of the Tier 1 and 2 models should be limited to about 2 years in length and recalibrated after this time. Strategy evaluation based on post hoc assessment through the intervention models can be carried out at any time post strategy implementation and will give more robust estimates of strategy effects, hence why this analysis has been recommended in the framework.

Practical application of the Tier 3 evaluation models shows their ability to articulate relative effects of key program activities, feedback from which can allow fine tuning of a road safety strategy to maximize performance and ensure set objectives are met. This level of evaluation can provide information about successful elements of a strategy without the need for full formal evaluation. Formulation and application of the Tier 3 evaluation models also needs to be carried out with great care by trained experts to ensure robust results. There are also limitations on the types of factors that can be considered in Tier 3 level models, generally being restricted to factors that have natural inherent variation over time that can be related to variations in the outcome measures being modeled. Factors with little monthly variation, generally those with high internal inertia such as changes in the profile of a vehicle fleet, are not suitable for consideration in Tier 3 type models and will need be investigated using other techniques. The success of the Tier 3 level evaluation will be limited by the availability and quality of data on road safety program outputs and may also be hampered by lack of a-priori knowledge of the most appropriate measure of output from a road safety program. As noted, the Tier 3 evaluation is not a substitute for formal and rigorous evaluation of specific road safety programs which remain an integral part of the evaluation framework in Tier 4.

Developing a road safety strategy evaluation framework such as the one proposed here has a number of key advantages. If formulated at the beginning of a road safety strategy or in parallel with development of the strategy, it provides a forward looking approach in on which to systematically plan and resource evaluation of the strategy. In doing so it also provides impetus to ensure the appropriate data are collected to facilitate the best possible evaluations and offers the opportunity to engage specialist evaluators at an early stage of the strategy to give advice on data collection requirements.

Test running the evaluation framework also helped identify suggested timings for each tier of evaluation activity. Tier 1 & 2 pre-hoc evaluation could be first carried out at strategy commencement and recalibrated after each action plan, or about every 2 years. Actual strategy progress against the projection would be monitored continuously between calibrations. Tier 1 and 2 post-hoc evaluation could be scheduled at the end of each action plan along with Tier 3 evaluation at the level of detail allowed by the available data. Specific program evaluation under Tier 4 would be carried out as required during the strategy with timing dependent on implementation time of each program, size of the program and availability of sufficient data for rigorous evaluation. Key programs for Tier 4 evaluation could be flagged at commencement of the strategy.

Conclusion

This study defines a framework for comprehensive evaluation of road safety strategies. Based on the GOSPA framework it defines a hierarchy of evaluation effort that assess strategy outcome effectiveness against the broad objectives of the strategy at the highest level right down to the effectiveness of specific programs carried out under the strategy at the finest detail. It provides a forward-looking basis for planning overall evaluation effort by road safety agencies that will assist in planning the specification of evaluation activities, supporting data collection and allocation of resources to facilitate effective and comprehensive road safety strategy evaluation.

Acknowledgement

This study was funded by Queensland Department of Transport and Main Roads (formerly Queensland Transport).

References

1. Johnston, I., *A brief review of Victoria's road safety outcomes and strategies.* Report to Department of Treasury and Finance. 2001, Monash University Accident Research Centre: Melbourne, Australia.
2. Cameron, M.H., *A method of evaluation based on a strategic planning framework* in *Proceedings Australasian Road Safety Research Policing and Education Conference.* 1999: Canberra, Australia. p. 801-08.
3. Haaland, K. and J. Odeck, *Economic evaluation methods for road projects in member countries: Compiled information on existing methodologies.* 1998, Norwegian Public Roads Administration, Committee of PIARC: Oslo, Norway.
4. Harvey, A. and J. Durbin, *The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling. Journal of the Royal Statistical Society. Series A (General)*, 1986 **149**: p. 187-227.
5. Harvey, A., *Forecasting, Structural Time Series Models and the Kalman Filter*. 1989: Cambridge University Press.
6. Frith, W.J. *Modest sample before and after studies of change to the road environment*. in *Institute of Engineers (New Zealand)*. 1987.
7. Leggett, L.M.W., *Using police enforcement to prevent road crashes: the randomised scheduled management system*, in *Crime Prevention Studies*. 1997, Wiley: New York, U.S.A.
8. Montgomery, D., *Introduction to Statistical Quality Control*. 2nd ed. 1991: John Wiley & Sons.
9. Newstead, S., I. Bobevski, S. Hosking, and M. Cameron, *Evaluation of the Queensland Road Safety Initiatives Package*. Report No. 272. 2007, Melbourne, Australia: Monash University Accident Research Centre.
10. Hosking, S., S. Newstead, E. Hoareau, and A. Delaney, *An evaluation of the 50km/h default speed limit in regional Queensland*. Report No. 265. 2007, Melbourne, Australia: Monash University Accident Research Centre.
11. Hoareau, E., S. Newstead, P. Oxley, and M. Cameron, *An evaluation of 50 km/h speed limits in South East Queensland*. Report No. 264. 2007, Melbourne, Australia: Monash University Accident Research Centre.
12. Newstead, S. and M. Cameron, *Evaluation of the crash effects of the Queensland speed camera program.* Report No. 204. 2003, Melbourne, Australia: Monash University Accident Research Centre.

13. Newstead, S., M. Cameron, and M. Leggett, *The crash reduction effectiveness of a network-wide traffic police deployment system.* Accident Analysis & Prevention, 2001. **33**(3): p. 393-406.