Risk assessment of rural intersections based on predictive modelling

Chris Jurewicz[a]

[a]Transport Accident Commission, Victoria, Australia

Abstract

Rural intersections present unique road safety issue: they are numerous and typically have low crash counts. Hence, crash history can be a poor predictor of future safety performance. Given that 10% of all Victorian severe crash injuries occur at rural intersections, their effective prioritisation for improvement is a priority.

This paper explores prioritisation of rural intersections based on severe injury estimates using three alternative approaches. The findings show that predictive models based on traffic flow, road hierarchy, intersection type, and maximum speed limit optimise both prioritisation logic and ease of applicability. These findings may be useful in development of future safety infrastructure investment programs.

Background

This paper explores prioritisation of rural intersections based on severe injury estimates using three alternative approaches.

Rural intersections accounted for 10% of all fatal and serious injuries (FSI) in Victoria, based on 2011-16 VicRoads crash data. Hence, this road element is a priority for Safe System treatments. However, rural intersections present a specific road safety issue – prioritisation of sites for potential improvement is difficult using historical crash data. The main issues are:

- Rural intersections are numerous – 7,786 were identified in VicRoads road data registers.
- Many had zero FSI count in the past ten years (21%).
- Where severe crashes were recorded, the FSI counts were low, 1-3 in ten years (48%).

Comparing sites with zero or low count data is difficult. Low crash counts become categorical in nature, with many different intersections in each category (e.g. 1 crash per site). Distribution of such low count data is highly random, and may not be due to some inherent design or operational flaw requiring an improvement. Hence, prioritizing rural intersections based on low crash counts may not be useful for identifying the need for treatments and for funding. A method which reliably identifies treatable risk factors associated with poor safety performance would be more useful.

Safe System infrastructure is delivered via transformative solutions, i.e. such that no future safety treatment is needed to minimise FSIs. If this is not viable, then cost-effective supporting solutions can be planned. It is then assumed that some future improvements will transform the location towards Safe System operation. In this context, VicRoads used the following criteria to develop an appropriate method for prioritising rural intersections for treatment:

1. Rank intersections based on risk of FSI, easily repeatable and updateable.
2. Quickly and easily estimate future FSIs at sites with and without significant high severity crash histories.
3. Understand the scale of existing FSI risk and potential for improvement for each intersection and for each intersection type.

The second and third objectives included a 'reactive' vs. 'proactive' consideration. Reactive approach models future safety performance mostly on crash history. Proactive approach considers presence and effect of known risk factors.

Methodology

The work presented in this paper builds on previous work undertaken by VicRoads and its consultants to create a data base of all Victorian intersections. Intersection data consisted of intersection type, urban / rural environment, number of legs, and traffic flows on each leg (AADT) where available. The abstract's objective was to present approaches rather than details of numerical methods. These will be presented in a full paper.

The intersection data was then geographically matched with crash injury data over a 10-year period (2006-2015). This new variable (FSI 10yr Weighted) was based on 2/3 latest five years & 1/3 previous five years. The main aim of this approach was to partially overcome the random variation due to low count numbers. This variable became the historical crash-based measure of severe injury for each intersection.

Different prioritisation approaches were tested against the objectives 1, 2 and 3 as described below.

Results

In general, prioritisation of intersections is based on their level of risk associated with FSI. Sites are prioritised on estimates of equivalent FSIs for a 10 –year period based on model input variables. The three compared methods vary from one which is reactive to crash history only, a hybrid which considers all variables including crash history, and a proactive method which does not require crash history to rank sites.

HRI Method - reactive

The first method tested was based on the approach described in the *High-risk Intersections* (HRI) guideline (New Zealand Transport Agency 2013). The HRI method uses a severity factor to convert recorded all-injury crashes for each site to a 10-year FSI equivalent estimate. The severity factors were based on analysis of all intersection data (urban & rural) and calculated for each speed limit and DCA group. This heuristic model indirectly accounts for speed and crash type, but only at intersections where crashes occurred. The FSI 10-year equivalent values (model) for each intersection were compared with observed values (10-year FSI Weighted) in Figure 1. The data scatter suggests that the model did not fit the historical data well.

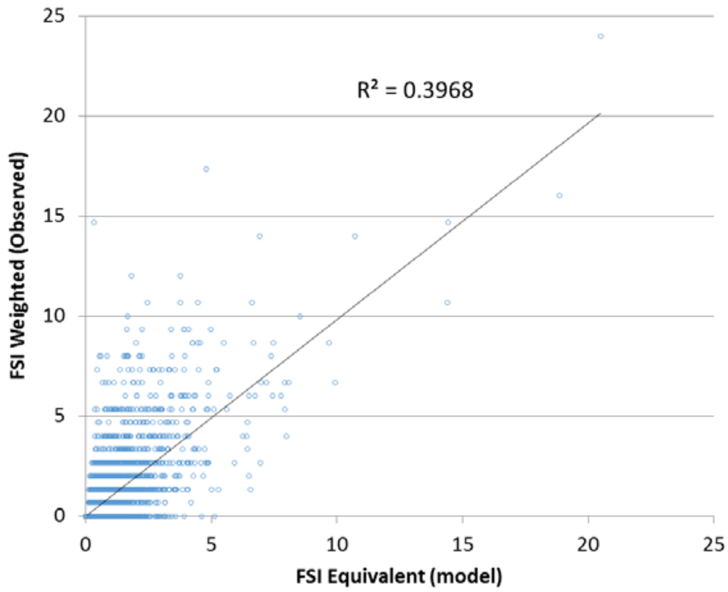

Figure 1. HRI method's FSI Equivalent (model) vs. FSI Weighted (observed) data

Table 1 tests this method against the intersection prioritisation objectives.

Table 1. HRI method vs. prioritisation objectives

Objectives		Assessment
1	Rank on FSI Easily repeatable and updateable	Must have casualty crashes recorded or all zero sites will be ranked the same. Can differentiate sites with same number of casualty crashes by crash type mix and by different speed limits. The model is very inaccurate and does not reflect typical performance (poor model fit, Figure 1). Overall, ranking based on the model is logical but not accurate for sites with low crash counts (i.e. the majority). Repeatable and updateable, as the calculation is simple once severity factors are available.
2	Quick and easy estimate of future FSI with and without crash problem.	This method works only for sites with casualty crashes recorded in the past 10 years. This is a limitation.
3	Estimate potential for improvement at each site and intersection type.	Due to the above limitations potential for improvement cannot be inferred from the model variables for intersections with nil or low crash counts. There was no distinction of intersection types.

Neural Network – hybrid reactive / proactive

Using the same data set described in the Methodology, a Neural Network model was developed for sites with recorded crash data in the past 10 years. Neural network is a type of machine learning technique. A model is trained to predict outputs using a large set of input and output data. Neural network techniques have been gaining use in crash frequency and severity estimation in recent years. This type of modelling can outperform conventional GLM and binary logistic models in some applications (e.g. Zang et al. 2016, Li et al. 2018).

Different input data variables were tested for importance in predicting the outcome (10-year FSI Equivalent). The most important predictor was the number of vehicle occupant injuries in the past

10-years. Intersection type did not show as important, nor did any other variables. Traffic flow data and other attributes were missing at many sites which affected their importance as a predictor.

In essence, this was hybrid reactive model as it turned out to be based on crash history, after proactively considering other variables. Still, Figure 2 shows that the model fit was much better for the Neural Network than HRI Method. Table 2 shows that this method suffered from some of the same performance issues as the HRI Method.

Table 2. Neural Network vs. prioritisation objectives

Objectives		**Assessment**
1	Rank on FSI Easily repeatable and updateable	Same logic issues as HRI Method but a better model. It is thus more useful for ranking but not ideal. Repeatable as the online model tool can batch-evaluate new sites. Model update or calibration can be done with new data.
2	Quick and easy estimate of future FSI with and without crash problem.	This method works well only for sites with casualty crashes recorded in the past 10 years. This is a limitation.
3	Estimate potential for improvement at each site and intersection type.	Due to the above limitations potential for improvement cannot be inferred from the model variables. Intersection type was not an important predictor variable, compared with injury history.

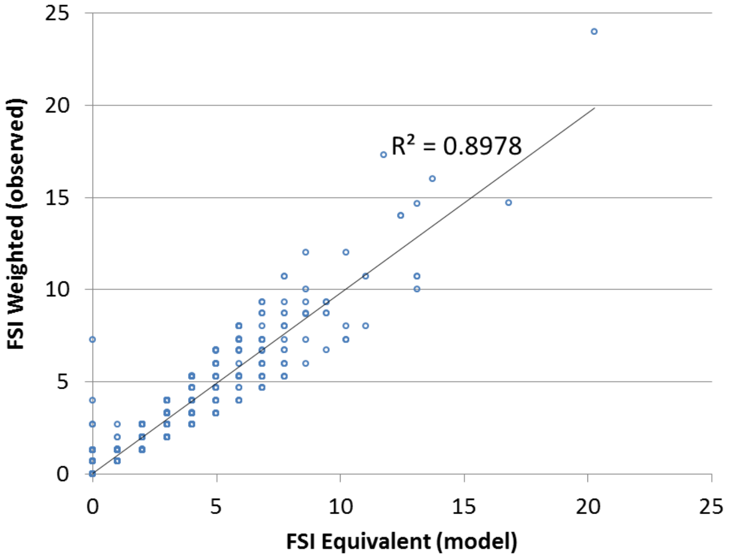

Figure 2. Neural Network FSI Equivalent (model) vs. FSI Weighted (observed) data

Random Forest Ensemble - proactive

A more proactive approach was taken in developing this model: it was based on linking known intersection attributes to their FSI Weighted history performance (10 years). No crash history variable was included to avoid the logical limitations of the previous two methods.

Random Forest Ensemble is a machine learning technique for predictive modelling. It is based on multiple decision trees estimating an outcome from known inputs (e.g. Pham et al 2010, Shi & Abdel-Aty 2015). This technique was selected from several alternative methods applied to the problem, as it produced the best predictive performance when validated against independent data.

The Ensemble identified several important FSI predictors: ‘Sum of AADT’ on all legs, ‘Intersection Type, ‘State/local’ (hierarchy of intersecting roads: state/state, state/local, etc.), and ‘Max Speed Limit’. Figure 3 shows that the model fit was not as good as the previous model, but still good. This is offset by more logical input variables which allow differentiation between sites with the same crash history (or lack of).

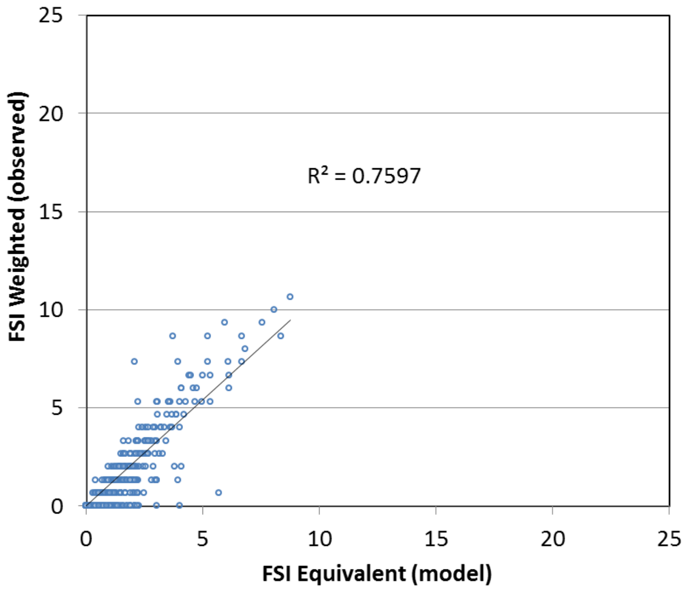

Figure 3. Random Forest Ensemble FSI Equivalent (model) vs. FSI Weighted (observed) data

This approach is consistent with the best practice in predictive risk estimation for intersections, which is typically proactive (e.g. Abdel-Aty & Haleem 2011, Turner et al. 2012). One model limitation was that only intersections with mostly complete attribute data had to be selected from the data set. Thus the model could be biased towards sites with good data (e.g. state/state intersections, subject to past investigations, etc.). Consequently, the model can only estimate FSI Equivalent values lower than 12 per 5 years (99.9% of all sites in the full data set). More complete future data sets will overcome all of these limitations. Table 3 assesses the model vs. the prioritisation objectives.

Table 3. Random Forest Ensemble vs. prioritisation objectives

Objectives		**Assessment**
1	Rank on FSI	All sites will be differentiated and ranked by relevant risk attributes, not crash history. The model is sufficiently accurate to provide confidence in results reflecting typical performance (good model fit, Figure 3). Overall, ranking based on the model would be logical and reasonably accurate.
	Easily repeatable and updateable	Repeatable as the online model tool can batch-evaluate new sites. Model update or calibration can be done with new data.
2	Quick and easy estimate of future FSI with and without crash problem.	Yes, estimates decimal FSI values for all intersections with 10-year FSI Weighted count under 12 (99.9% of sites in the data). Requires AADT data/estimate for all intersection legs, intersection type, basic state/local road hierarchies and the maximum speed limit.
3	Estimate potential for improvement at each site and intersection type.	Given the model uses risk attributes, improvement potential can be estimated, e.g. change from priority intersection type to roundabout.

Discussion and Conclusions

The analysis demonstrated the how logic and applicability objectives for rural intersection prioritisation can be met with reactive (crash-based), hybrid and proactive (predictive) risk measurement approaches. Comparison of prioritised lists of intersections could be part of a future exercise to test for meaningful difference.

Reactive methods based on historical injury crash data could not accurately differentiate between all intersections, although were easier to apply as only crash data was required. These methods also provided limited inputs into future improvement potential of intersections.

The proactive approach, based on predictive estimation of 10-year FSI Equivalent values, showed the best logic and applicability match. It differentiated all sites using inputs of Sum of AADT, intersection type, intersecting road hierarchies and maximum speed limit. This approach allowed prioritising rural intersections with zero and low injury counts. Applicability was somewhat reduced by the requirement for basic data for each intersection. AADT measurement or estimation across the rural road network may be necessary for future prioritisation.

References

New Zealand Transport Agency. (2013). *High-risk intersections guide*. Wellington, New Zealand: New Zealand Transport Agency.

Zeng, Q., Huang, H., Pin, X., Wong S.C. (2016). Modelling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. *Analytic Methods in Accident Research, 10, 12-25.*

Li, Y., Ma, D., Zhu, M., Zeng, Z. & Wang, Y. (2018). Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network. *Accident Analysis and Prevention, 111, 354–363.*

Pham, M.H., Bhaskar, A., & Dumont A.G. (2010). Random forest models for identifying motorway Rear-End Crash Risks using disaggregate data. *13th International IEEE Conference on Intelligent Transportation Systems, 19-22 September 2010, Funchal, Portugal.*

Shi, Q., Abdel-Aty, M. (2015) Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies, Volume 58, Part B, 380-394.*

Abdel-Aty, M. & Haleem, K. (2011). Analyzing angle crashes at unsignalized intersections using machine learning techniques. *Accident Analysis and Prevention, 43, 461–470.*

Turner, S., Singh, R. & Nates, G. (2012). *The next generation of rural road crash prediction models: final report*. Research Report 509. Wellington, New Zealand: New Zealand Transport Agency.