Applying Cluster Analysis to Validate a High Risk Young Driver Model: Implications for Tailored Road Safety Intervention

Bridie Scott-Parker[a,b,c], Leanne Stokes[d], Stuart Gardner[d], Megan Cawkwell[e], Matthew Wilson[f], Shane Panoho[f], Sherryn Klump[f]

[a] Adolescent Risk Research Unit, Sunshine Coast Mind and Neuroscience – Thompson Institute, [b] School of Social Sciences, Faculty of Arts, Business, and Law, University of the Sunshine Coast, [c] Sustainability Research Centre, University of the Sunshine Coast, [d] Transport and Main Roads, [e] Sunshine Coast Council, [f] Queensland Police Service

Abstract

This project explored the feasibility and efficacy of a model that could be used to identify high risk young drivers (HRYD). The present findings pertain to the final stage, which was to assess the predictive ability of the HRYD model. The two-step cluster analysis method was employed with de-identified Police records of 2,973 Sunshine Coast residents aged 17-24 years. The four clusters in the final solution aligned relatively consistently with the HRYD clusters developed during the earlier phases (literature review, focus groups) of this pioneering project. The HRYD model can be used to guide intervention targeting high risk youth behaviour.

Background

In 2014, funding was secured for a large-scale multi-disciplinary project to explore the feasibility and efficacy of a model to identify HRYD, a group of young drivers who exhibit particularly risky driving styles. This project commenced with a literature review of profiling of offending youth, including HRYD, and a qualitative study into the behaviours and characteristics that can identify HRYD. Subsequently these findings were integrated within the theoretical framework of systems thinking which asserts that young driver safety emerges as a result of a wealth of interactions between a multitude of key stakeholders, including police, and government licensing authorities (Scott-Parker, Goode, & Salmon, 2015; Scott-Parker, Goode, Salmon, & Senserrick, 2016), culminating in a HRYD model featuring on-road (eg, offences), and other (eg, shoplifting, risky peers) behaviours/characteristics. The final stage of this project, reported here, assessed the predictive ability of the HRYD model.

Method

Queensland Police Service provided de-identified data of all driving and non-driving-related occurrences (eg, fingerprinting for security licences) pertaining to Sunshine Coast residents aged 17-24 years (as of 30 June 2016), via records from age 14 years. Following an extensive cleaning process,[1] records of 2,973 individuals remained (98% of the original data). The two-step cluster analysis method was employed, with guidance from Hair, Black, Babin and Anderson's (2010) six-stage model-building approach. The cluster variate for the final solution comprised 10 variables pertaining to traffic crashes, drink/drug-driving, registration, dangerous/careless driving, driving causing death/grievous bodily harm, and driving while unlicensed/suspended/disqualified offences.

Results

The cluster quality[2] was 'Good' (0.6 out of 1.0). Four clusters were identified in the final solution (see Figure 1, Table 1). While individuals in both Clusters-2 and -4 were more likely to be male, broadly speaking Cluster-2 (HRYD) was distinct from Cluster-4; individuals in Cluster-2 having

---

[1] E.g., to remove rows of redundant data and occurrences where the individual was not classified as an "offender", "charged", and/or "driver" etc.

[2] A measure of the cohesion and separation of the final solution.

longer offence histories which contained significantly more driving and non-driving related offences in most instances (compared to any other cluster). Cluster-4, the smallest cluster, is distinct from Cluster-2 in that these individuals contain records that show few offences other than those pertaining to substance misuse. Conversely, Clusters-1 and -3 exhibited much more overlap. Individuals in Cluster-1 had records with few offences (often no driving offences) and little involvement in crashes other than those without injury. Cluster-3 individuals exhibited significantly more involvement in crashes with injury. Relatedly, individuals in Cluster 3-also exhibited more driving-related offences than Cluster-1.

Cluster-2: HRYD

- Higher rates of most offences (including driving- and non-driving related offences)
- Higher rates of hit and run crashes
- Contains all fatal crashes

Cluster-4: Substance users

- Highest rates of drink- and drug-driving
- Few other driving- and non-driving related offences (except offences for driving without a valid licence)

Cluster-1: Non-injury crashers

- Highest rates of crashes without injury
- Lowest rates of all driving-related occurrences and non-driving offences
- No driving-related offences

Cluster-3: Injury crashers

- Highest rates of crashes with injury
- Few other driving- and non-driving related offences
- Distinguished from cluster 1 by low rates of driving-related offences
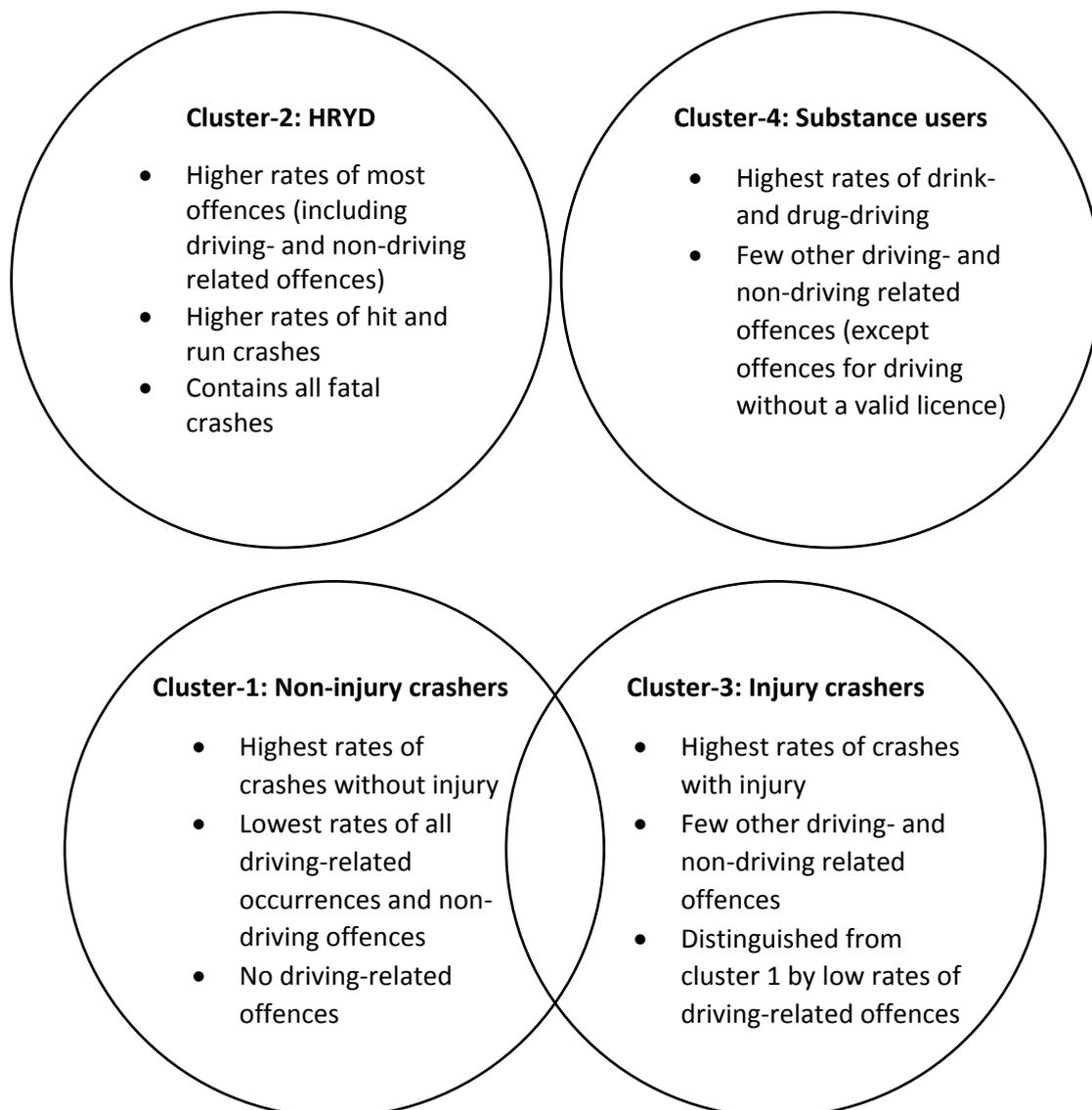
Figure 1. Cluster summary, illustrating that while Clusters 2 and 4 are generally distinct from all other Clusters, Clusters 1 and 3 are not always distinct

Conclusions

The findings of the cluster analysis as they pertain to HRYD specifically were found to align relatively consistently with the characteristics of a HRYD model developed during the earlier phases (literature review, focus groups) of this pioneering project, validating the findings of these earlier research stages. Thus this methodologically-rigorous multi-stage process has provided a theoretically-sound model that can be tested in other youth offender populations. Moreover, the model can be used to guide the development, application, and evaluation of intervention targeting a variety of high risk

youth behaviours, including on-road driving behaviours. Such targeted intervention (eg., Cluster 4 youth are likely to benefit most from harm minimisation interventions targeting substance use/abuse difficulties; the characteristics of youth in Cluster 1 suggests that early, multi-faceted intervention is required) is likely to result in the improvement of the health and wellbeing not only of the targeted youth themselves, but also all with whom the youth shares the road.

1   *Table 1. Summary of the profiles of the four clusters*

Characteristic/factor	Cluster 1: Non-injury crashers	Cluster 2: HRYD	Cluster 3: Injury crashers	Cluster 4: Substance users
% of sample (*n*)[a]	21.4 (636)	40.1 (1,191)	23.8 (709)	14.7 (437)
% of 17-24 years population [b]	2.5	4.6	2.8	1.7
Gender [c]	More likely to be female* (59.9%, *n*=381 male)	More likely to be male* (80.0%, *n*=953 male)	More likely to be female* (53.2%, *n*=377 male)	More likely to be male* (77.8%, *n*=340 male)
Age at offending (years)	Older at first offence ($\overline{x}$=18.2), younger at last offence ($\overline{x}$=18.9)*	Younger at first offence ($\overline{x}$=17.3), older at last offence ($\overline{x}$=19.8)*	Older at first offence ($\overline{x}$=18.1), younger at last offence ($\overline{x}$=19.2)*	Older at first offence ($\overline{x}$=18.3), younger at last offence ($\overline{x}$=19.6)*
Non-driving offences	Few	Highest	Few, not as few as Cluster 1	Generally few, except for substance-related**
Crashes	Non-injury only	Fatal, hit and run, injury, non-injury	Non-injury and injury	None
Driving offences	None	Many	Few	Generally few, except for substance-related
Dangerous/careless driving	None	Highest	Few	None
Licence-related	None	Highest	Few, not as few as Cluster 4	Few
Registration-related	None	Highest	Few	None
Substance-related	None	Many	Few	Highest

[a] Sample population *n*=2,973
[b] 17-24 years population on the Sunshine Coast, data compiled for postcodes present in the dataset only, *n*=25,770 (ABS, 2016).
[c] Compared to the overall sample (of which 69.0%, *n*=2051 were male).
* Chi square analyses, *p*<0.001.
** Krukal-Wallis pairwise cluster comparisons *p*<.001 for all cluster comparisons except between clusters 1 and 3.

2

3 **References**

Australian Bureau of Statistics (2016). Census table builder: 2011 Census of Population and Housing, POA by Age in Single Years (AGEP). Retrieved from http://www.abs.gov.au/websitedbs/ censushome.nsf/home/tablebuilder

Hair, J. F., Jr., Black, W. C., Babin, B. J., Anderson, R. E. (2010). *Multivariate data analysis*. 7th ed. New Jersey: Pearson Education, Inc.

Scott-Parker, B., Goode, N., Salmon, P., Senserrick, T. (2016). Knowing me knowing you: key players and their interactions within the young driver road safety system. *Safety Science, 88,* 88-96.

Scott-Parker, B., Goode, N., & Salmon, P. (2015). The driver, the road, the rules.....and the rest? A systems-based approach to young driver road safety. *Accident Analysis and Prevention, 74,* 297-305.