LEVERAGING THE DIFFERENCES IN CRASH PROFILES OF FLEET DRIVERS USING MOTOR INSURANCE CLAIMS DATA

Mr Phil Darby
Edinburgh Napier University
Dr Will Murray
Interactive Driving Systems
Professor Stephen Ison
Dr Mohammed A Quddus
Loughborough University

ABSTRACT

Employees driving for work have a proportionately high risk of involvement in a fatal or serious road traffic collision. For example, it has been estimated that 25% of UK road accidents involve someone driving for work. Across the EU, 34% of work fatalities involve road traffic or transport accidents. These collisions are not confined to large trucks but include many smaller vehicles where driving is secondary to the employees' main tasks. This gives relevance to the study of improving fleet safety. Increasingly fleet risk management has focused on safety and the behaviours of drivers rather than asset management and cost controls. Studies have shown that changing behaviour in both drivers and the rest of the organisation can make large improvements in safety. This study reports on the analysis of detailed insurance claims data from a large UK motor fleet of approximately 40,000 vehicles, since 2001. A one size fits all approach cannot be expected to work with so many drivers in such a variety of roles. Although each worker is an individual, he or she share patterns of behaviour with those of fellow workers. Rather than treating employees as an amorphous mass they are clustered into homogeneous groups using multivariate statistical techniques based on driver claim histories. The influence of interventions on the behaviour of these groups will contribute to successful safety improvement strategies for each group. The performance of these groups has been studied and interventions successfully applied. Some of these are reported in the paper. The company has had a good track record in applying a range of strategies to produce significant performance improvements. Latterly the rate of improvement has shown signs of slowing so a new view of the drivers and the system in which they operate is needed. One such approach could be a reconsideration of the segmentation scheme based on the clustering approach.

1 INTRODUCTION

Driving for work is increasingly recognised as a major health and safety risk and represents a significant portion of both worker fatalities and road fatalities. Those who drive for work can be employed in a wide range of roles from those are full time drivers to those who run an infrequent office errand or others who simply commute. A many employees drive high mileages while carrying out such jobs as a managers or performing technical tasks at customer sites. Much research and policy attention has focused on large and heavy commercial vehicles in contrast other frequently used fleet vehicle types which often constitute a major part of an organisation's fleet.

Concerned companies understand the high risks faced by their employees so many have put systems in place to monitor and reduce them. This is achieved by addressing broad cultural approaches and specific counter measures (Murray, et al., 2009). The choice of issues addressed comes from domain knowledge and by drilling down to find similar types but this does not guarantee homogeneity since subgroups of these segments may remain hidden. Even when investigating one vehicle type other factors are important. Savolainen and Mannering (2007) highlighted the need to better understand interrelated risk factors in their study of motorcyclists' injuries.

This paper reviews the fleet safety strategy and performance of a company using their insurance claims as a measure of performance. The company's safety programme is demonstrated as being very successful. The current rate of improvement does not match their rapid initial success. A new approach to exploring the data is proposed and demonstrated using clustering techniques. The issue of having mostly categorical data is overcome newer using techniques widely used in field such as marketing. The work concludes by indicating how this approach can identify previously hidden data trends then explains how the approach can be used to update policy and practice.

2 BACKGROUND

The subject of the study is a multinational company with 100,000 employees many of whom could be expected to drive for work. Approximately 60,000 of the employees are classified as drivers. Most of the drivers have a technical, managerial or sales role but use a vehicle to travel between jobs. The company has a varied fleet of around 40,000 vehicles two thirds of which are 'white vans'. The company also uses hired vehicles, and some employees drive their own vehicle on work business. It has identified traffic collisions as the largest health and safety risk leading to several deaths and serious injuries among company employees

Employee crash involvement has been improved by addressing risk factors on many fronts and applying appropriate interventions. The company has introduced 127 interventions to improve fleet safety using the Haddon framework to assess risks and reduce their impact (Wallington, 2009; Darby et al. 2010). This creates 15 possible intervention categories by considering five topics (management culture, journey, road or site environment, people, vehicle and society or community across three time periods (before the drive or crash, at the scene and after the drive or crash). The management culture of safety is exemplified by the appointment of a board level champion, giving emergency support to drivers and systematic investigation of crashes. Examples from the journey dimension include questioning the need to travel, journey planning and route selection, while including safety while selecting, recruiting and training staff indicates how the people aspect of the matrix is addressed. No single item has been identified as a silver bullet but the basket of approaches has contributed collectively to the overall improved performance.

An important element of the program is the continuous risk assessment of drivers. Under this regime employees are first identified as potential drivers. Only non-drivers are exempt from the programme even those rarely driving on business must engage. Only 10% of employees are classed as non drivers. Drivers are made aware of data protection issues and how this data will be gathered stored and used. Data security is assured though absolute adherence to the strongest possible protocols which have been agreed with labour organisations. Employees have to agree that their information is used in this way. Next drivers must pledge to drive safely to recognise their role in safety before undertaking online risk profiling and defensive driving assessments. The driver's level of crash risk is determined from both these combined with knowledge of the type of driving they undertake Approximately 25% of those assessed experience low risk, 73% medium risk and 2% high risk of making a claim.

Counselling by the individual's line manager takes place for those identified at elevated risk. Drivers at little risk receive no mandatory training. Those at middle levels of risk are referred to attitude and behaviour training. Only those at the highest levels of risk or have reached a threshold of incidents are referred to in vehicle training. 76% of these sessions focus on 'safed' (defensive + eco driving), 11% on low speed manoeuvring, 8% are related to induction, (e.g. apprentices) and 5% on high mileage drivers. The differing sessions aim to target training towards varying aspects of risk that drivers face.

The gains made by using approach can be seen in the company's claims reduction history. The nine year trends in the number of claims (excluding glass) together with STATS19 data of the number of accidents reported in Britain where someone was killed or seriously injured (KSI) are displayed in Figure 1. Data showed that claims reduced by 6.5% per year and KSI accidents by 3.9% per year on average. Some monthly claim totals have been removed from the analysis such as the end points and two months in mid 2006 when transitional reporting anomalies were rectified. The trend in KSI accidents has a clear seasonal component which is absent from the claims data since work based exposure does not reduce in winter.

Three claims handlers were used during this time with each using a different reporting system. The plot points out the duration of each handler. Analysis has focused on the final claims handler (since mid 2005) to remove possible confounding of the analysis. A second model of claims and seasonally adjusted KSI accidents since mid 2006 indicates that both are reducing by 3.3% per year over the last period. The company's performance is highly creditable since claims mostly consist of damage only incidents while road safety policy is directed towards reducing personal injury. The motivation for further work in this paper is the recognition that there may be future reductions in the rate of improvement and attempts to identify opportunities for improvement.

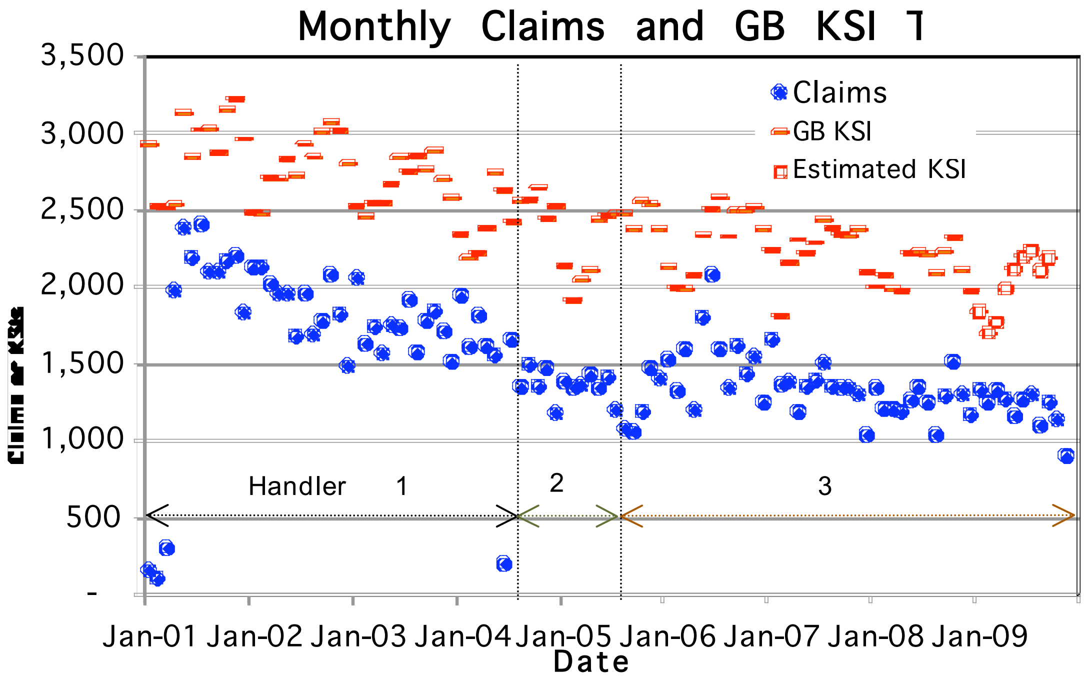


Figure 1 Trend in crash claims and GB KSI accidents. (Estimated KSI values derived from provisional quarterly results)

The current approach to safety analysis has been to identify 'one-dimensional' trends, e.g. is a certain claim type increasing or decreasing. Opportunity exists to analyse data using multidimensional techniques of multivariate statistical analysis such as clustering that are common in other fields (Wedel and Kamakura, 1998).

3 BACKGROUND TO CLUSTERING METHODOLOGY

3.1 Clustering and Segmentation

Clustering is a statistical approach to forming groups which "*provides a means for assessing dimensionality, identifying outliers and suggesting interesting hypotheses concerning relationships*" (Johnson and Wichern, 2002). Clusters are defined by the data rather than being imposed by external factors. Thirteen types of clustering algorithm have been identified in a literature (Xu and Wunsch, 2005). These include hierarchical clustering, neural networks and mixture density approaches which are used in this study.

Segmentation has a history of use in marketing (Smith, 1956; Wedel and Kamakura, 1998) to identify homogeneous groups of customers whose wants and needs can be better served by different products or services. Other fields using clustering techniques include pattern recognition, web mining, genetics, geography and sociology (Xu and Wunsch, 2005). Uses in traffic safety have included risk taking in young drivers (Ulleberg, 2001) and characterisation of fatal crashes amongst older drivers (Skyving, et al., 2009).

3.2 Similarity of Observations

Clustering assumes that relevant similarity/dissimilarity measures can be produced and represents differences between objects, e.g. people, crashes, and companies. Analysis minimises the intra-group differences while maximising the inter-group differences. Statistical clustering has used Euclidean distance as a measure of dissimilarity using continuous data. Categorical data is more difficult to incorporate in the same way as it lacks a proportional scale. The number of matches between categorical variables has produced indices as a proxy for similarity, e.g. Jaccard index or Hamming distance (Johnson and Wichern, 2002).

3.3 Hierarchical Clustering and Latent Class Analysis

Putting the data onto a common scale by variable standardisation overcomes the problem of some variables dominating the analysis by virtue of their measurement scale. Near neighbours become linked into clusters using rules of separation (i.e. nearest neighbour) or by allocating variables in a way that minimises the within cluster variance relative to the between cluster variation (Ward’s method).

Model based clustering approaches have been developed over the past 20 years (Fraley and Raftery, 2002) including latent class analysis (LCA) indicated in general form in equation 1. The probability of the data (Y) given the model parameters (□) is given by the product of individual probabilities of cluster membership with k clusters using a prior belief of $P(C_z)$. Cluster membership (z) and parameters ($□_z$) are varied to maximise this probability for example using the Bayesian information criterion (BIC).

$$p(Y \mid \theta) = \sum_{z=1}^{K} P(C_z) \left[\prod_{j=1}^{m} p(y_j \mid C_z, \theta_z) \right] \qquad -1$$

4 CRASH CLASSIFICATION

4.1 Non Statistical Crash Classification

Road traffic incident analysis is frequently performed by addressing individual characteristics such as the vehicle or road type. This is an intuitive approach since crashes that share characteristics should be more homogeneous and could be treated uniformly when designing useful intervention. Examples of the approach are motor-cycles (Quddus, et al., 2002), trucks (Häkkänen and Summala, 2001; McCall and Horwitz, 2005) and cyclists (Stone and Broughton, 2003). This leads to decision tree like structures from drilling down to find interesting groups and does reduce road casualties. However, this does reduce group sample sizes and could miss events that share other characteristics as the search space is reduced.

Combinations of conditions have been studied with a view to better understand risks. Doherty, et al. (1998) reported the greatest crash involvement for 16-19 year old Ontario drivers at night-time when carrying passengers. Clarke, et al. (2006) reported relatively more accidents with male drivers negotiating curves on single carriageway rural roads at night-time. McCall and Horwitz (2005) studied truck accidents in Oregon between 1990 and 1997 and found males constituted the majority (80.7%), most were 35 years old or younger and had been in their job for less than one year. High death rates per vehicle mile in both younger and older drivers (Li, et al., 2003) resulted from excess crash involvement of younger drivers but greater frailty amongst older drivers. Motorcyclists involved in serious incidents were over represented as being to blame on bends and curves and as victims of right of way violations (Clarke, et al., 2007). Analysis of work-related crashes in France indicated relative risks varied markedly by gender, journey purpose (work, commuting or non-work) and occupational category (Charbotel, et al., 2001).

These studies indicate the need to identify groups involved in similar accidents in order to create separate models for each group.

4.2 Crash Clustering

Measures other than grid location can be incorporated into the calculation of similarity. Researchers created clusters with neighbourhood characteristics added to crash locations to improve the model accuracy by cluster (Levine, et al. 1995; Ng, et al. 2002) Variable standardisation overcomes scale imbalances. The approach is restricted to continuous data or by averaging categorical data over a grid to give the number within the grid or distance from the event location.

Other approaches to accident hot spots identification are spatial autocorrelation and kernel estimators (Flahaut, et al., 2003). Anderson (2009) used the kernel method to evaluate the effects of proximity to features such as cycle lanes, traffic lights and schools on accident hotspots. A general approach to clustering accidents applying LCA to both continuous and categorical data been demonstrated by Depaire, et al. (2008).

The approach of using the data to define clusters as used by Depaire et al. (2008) has been used here since similar crash related data is available. Clusters will be defined by the data to occur without reference to inputs from crash experts. The output would assist experts to analyse crash data.

Crash hotspots cannot be identified since no crash position information is available.

5 DATA

The data comprised of 72,777 incident claims over a 4½ year period from mid 2005 to the end of 2009. The claims range in severity from fatal crashes to vandalism and broken glass. A summary of the claim types is given in Table 1 below. Supplementary driver information is available through the outcomes of a driver assessment programme (Darby, et al., 2009) and limited demographic profiles using HR data. The age and gender of the claimants was missing from this dataset so demographic information from a single division was used to supplement the claims data. This reduced the sample size to just over 36,000. The proportion of claims in each category in the enhanced data did match those in the full data.

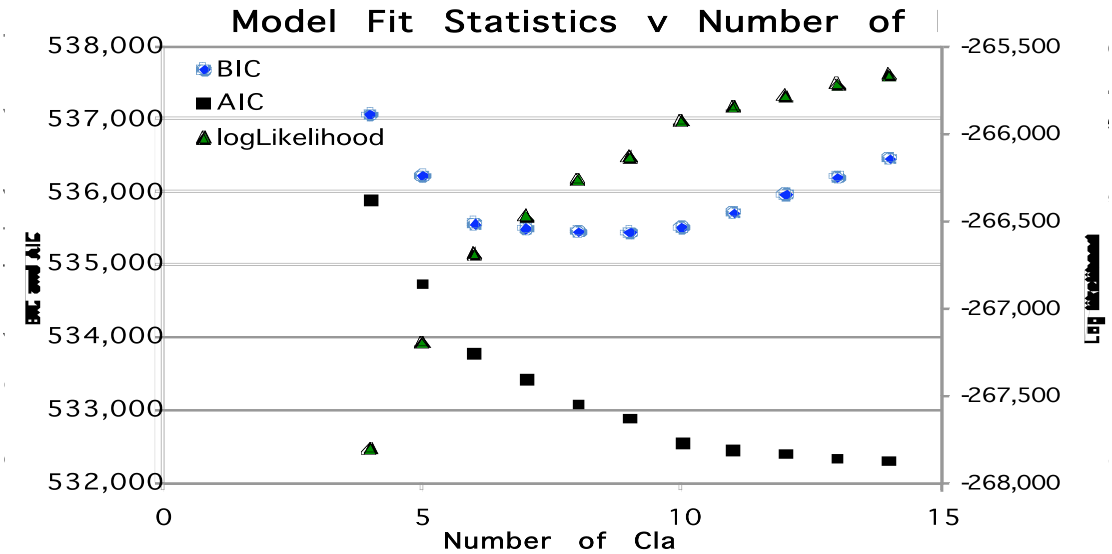
Table 1 Frequency of claims made by fleet drivers in the entire data set and that used for model building

Claim type	% in all	% in sample
Hit while parked	23.3	24.1
Other	21.4	23.2
Hit a stationary object	19.0	17.3
Crashed while driving	11.0	10.4
Fire, theft or vandalism	7.1	7.0
Rear ended	5.7	5.8
Hit by other while driving	5.4	5.2
Reversing or low speed manoeuvre	0.2	4.5
Crashed while turning	1.3	1.1
Incident happed previously	0.8	0.7
Information not available	0.1	0.3

Each claim recorded the date and time of the incident from which the season, weekend flag and 2 hour time slot were calculated. No location information was available such as that given in the STATS19 data base. A detailed vehicle description was available but these were condensed into car, light or heavy commercial vehicle. A set of claim types were available which were mapped onto 11 types such as ‘hit while parked’ to ‘hitting stationary objects’. A flag for fault or no fault was added to each claim by the claims handler after discussion with the reporter. Crash severity was available in both human terms and claims costs by source (e.g. damage, third party claims or cost of vehicle replacement).

6 ANALYSIS

Analysis closely followed the work of Depaire, et al. (2008) although this work used the poLCA[1] routine (Linzer and Lewis, 2007) written in the R statistical computing environment rather than Latent Gold (Statistical Innovations, 2008). Only categorical variables are allowed in poLCA but all variables were categorised and followed the reference coding scheme where comparative data was available.



times for larger models.

Figure 2 Change in BIC, AIC and log likelihood as classes are added to the model

6.1 Class characteristics

Since the objective of clustering is to reduce the homogeneity of the classes the profile of each class has been examined to class to class differences. Summaries of the nine classes are shown in Table 2 each labelled with A to I where Class-A was the largest class with 32.2% of the observations and Class-I is the smallest having only 2.0% of the observations. The characteristic profiles of each class are given along with the percents by category for all the data.

A brief description of each class is given in the lower section of Table 2. This was obtained by inspection of the class percentages compared to the overall value rather than using statistical analysis. The “at fault” variable is a key determinant of class membership. Three classes have most drivers ‘at fault’ while six have most drivers ‘not at fault’. The majority of injury claims are in Class- E in which 12.8% claims relate to an injury or fatality. Classes-F and I contain claims from the weekend while I contains the majority of night-time incidents. Classes A and B contain no cars while Classes-F and H contain mostly cars. Class-B and F relate to a large number of hit-while-parked claims while Class H contains a higher proportion of female drivers. Class-A contains younger drivers while Class C contains older drivers.

Table 2 upper Percent in each category by class
lower Outline description of each class

		Class									
		A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**All**
% in class		32.2	29.4	12.0	6.7	6.4	5.0	4.0	2.4	2.0	na
Fault		100.0	-	-	-	-	-	99.96	100.0	-	38.5
Injury or fatality		1.2	0.2	0.3	0.0	12.8	0.2	0.4	0.8	0.5	1.3
Weekend		5.7	4.4	2.1	12.1	5.5	21.1	6.2	19.9	55.8	7.7
Time	7-9	22.9	24.3	27.0	51.5	32.8	18.7	23.2	21.3	-	25.9
	10-12	30.5	33.0	34.7	13.6	16.7	25.2	32.4	23.0	15.1	29.0
	1-5pm	31.1	31.4	31.5	8.7	24.5	23.9	33.0	22.0	3.6	28.0
	4-6pm	13.1	10.3	5.5	8.0	23.1	22.7	8.3	20.9	4.0	11.3
	7-9pm	0.8	0.5	0.5	6.7	0.8	6.8	1.0	8.1	12.5	1.9
	Night	1.6	0.6	0.7	11.5	2.1	2.7	2.2	4.8	64.9	3.9
Vehicle	Car	-	-	13.4	2.0	5.3	99.4	-	100.0	14.4	9.9
	LCV	99.8	100.0	76.3	92.1	92.0	-	74.9	-	81.0	85.1
	HCV	0.2	0.0	10.3	5.9	2.7	0.6	25.1	-	4.6	5.0
Claim	HWP	-	50.6	30.6	32.6	10.7	50.3	-	-	27.9	23.3
	Other	3.4	36.0	54.9	-	-	26.7	11.0	8.3	11.8	21.4
	Stationary	48.5	-	-	-	-	-	52.2	46.6	-	19.0
	Driving fwd	27.4	1.4	1.7	1.8	5.8	2.2	19.6	27.3	1.5	11.0
	FTV	-	2.1	0.9	61.2	-	6.0	-	0.1	54.8	7.1
	R-ended	1.6	1.5	6.5	-	53.0	7.0	1.5	2.1	1.9	5.7
	Hit by Other	-	8.2	5.2	4.4	30.5	7.6	-	-	2.3	5.4
	Reversing	12.4	-	-	-	-	-	11.9	9.7	-	4.7
	Turning	4.1	-	-	-	-	-	1.4	3.5	-	1.3
Female		2.4	2.3	0.1	0.8	1.4	9.6	0.1	6.4	1.7	2.0

Age Group	18–21	0.1	0.1	-	-	0.2	-	-	-	-	0.1
	22–29	5.4	3.6	-	1.4	2.2	-	-	0.2	2.1	2.6
	30–39	18.9	17.6	2.0	7.6	11.2	4.1	2.0	3.2	6.3	11.4
	40–49	27.2	27.2	12.5	17.1	27.1	12.0	15.1	13.2	21.9	21.7
	50–59	28.7	34.0	35.3	39.0	37.0	44.0	40.5	41.0	36.6	34.7
	60–65	18.8	16.6	46.7	32.6	21.3	39.6	38.4	41.8	31.1	27.8
	65+	0.9	0.9	3.5	2.3	1.2	0.4	4.1	0.7	2.0	1.8

Outline description of classes

A At fault; No cars; Driving and reversing; Younger;
B No fault; Light commercial; Hit while parked or Other;
C No fault; More cars; Older
D No fault; Night & Morning; Light commercial; Hit while parked or other
E No fault; Mostly light commercial; Hit while parked or Fire theft & vandalism
F No fault; Weekend; Cars; Hit while parked; Females; Somewhat older;
G At fault; Minor, severe or fatal; Mostly LCVs; Driving or turning;
H At fault; Cars or LCVs; Hitting stationary objects, driving or reversing; Females; Older;
I No fault; Weekend & Night; Fire Theft or Vandalism;

6.2 Class trends

Since latent classes have been identified further analysis can be undertaken. The monthly trend of claims using the nine classes is shown in Figure 3 below. This corresponds to the time period in Figure 2 where data was gathered by claims handler 3. Class-A is the largest claim group which also showed a large decrease over the observed period. Claims in Class-B have remained constant but with high variability. Class-C shows the most interesting characteristics of any group as claims in this class had been fairly constant but a marked increase since mid 2008 is observable. The remaining Classes-D to I have remained at a low level over the time of the data.

It should be noted, however, that despite the small size of Class-E (6.4% of the total) this does contain a large proportion of injury incidents. There is some evidence that this class is also reducing in size.

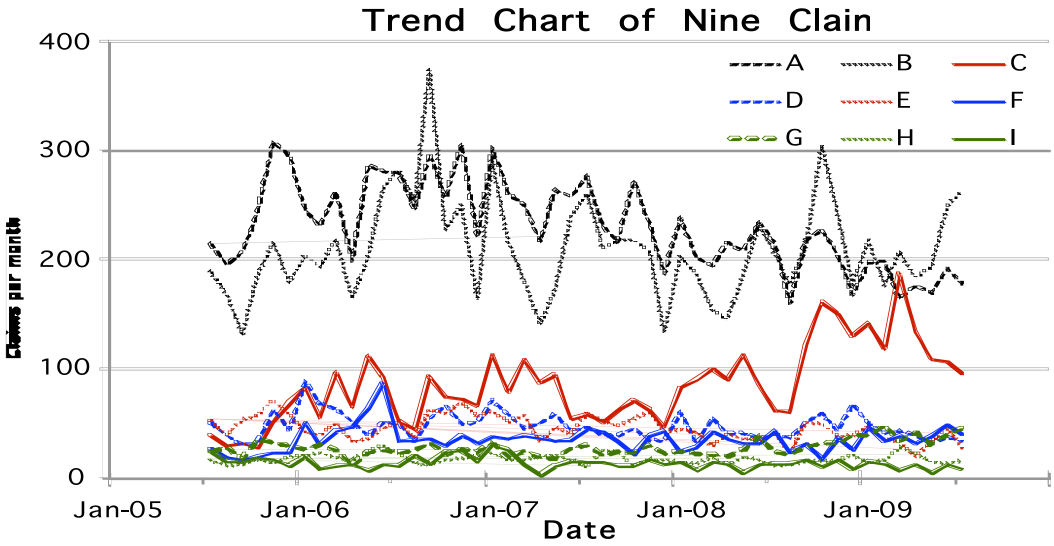


Figure 3 The trend of claims for each of the nine latent classes

Having identified these classes further analysis can be carried out. Class stability over time has not been investigated and that the class composition is constant. Data from earlier claims handlers was not used since it was felt that there could be a mismatch in category descriptors. Careful alignment of categories would be needed to accomplish this. Alternatively analysis could be restricted to data in which there is high confidence.

A major use of the groups would be to investigate the effectiveness of safety interventions on each of the classes. An example would be the use of driver training or assessment used as a predictor of class trends. This would determine if interventions were effective against claims of a certain class.

Injury and costs analysis should be carried out for each class in order to determine any possible differences in magnitude and direction dependences of other explanatory variables. Work by others (Savolainen, et al., 2007) has recognised the complexities of understanding behaviour within a single road user group.

7 CONCLUSIONS AND IMPLICATIONS FOR PRACTICE

The paper has demonstrated that the improvement in the safety of a large vehicle fleet outperforms that of a national benchmark. The approach to managing this success has been indicated with reference to the safety management with the training of at risk drivers as part of that approach. The reduction in the improvement rate has indicated that new approach may needed. The use of segmentation methods to find underlying groups within fleet vehicle claim data has been investigated as a possible way forward.

Latent class analysis proved to be useful in that it identified nine latent classes within the claims data. The concentration of factors into certain classes makes them both more recognisable and more homogeneous than the original data. These groups can enhance our understanding of driver behaviour and hence further improve performance. A key grouping variable was whether or not the driver was indicated as at fault further work in needed to determine if this is a trivial finding. Other factors are less trivial and offer insights. These results show that segmentation using LCA should be used more widely employing all possible data when investigating fleet safety.

Notes

[1] poCLA; Polytomous Latent Class Analysis

Bibliography

Anderson, T. K. (2009) Kernel density estimation and K-means clustering to profile road accident hotspots, **Accident Analysis and Prevention**, **41**(3), 359-364.

Charbotel, B., Chiron, M., Martin, J.L. and Bergeret, A. (2001) Work-related road accidents in France, **European Journal of Epidemiology**, **17**(8), 773-778.

Clarke, D. D., Ward, P., Bartle, C. and Truman, W. (2006) Young driver accidents in the UK: The influence of age, experience, and time of day, **Accident Analysis and Prevention**, **38**(5), 871-878.

Clarke, D. D., Ward, P., Bartle, C. and Truman, W. (2007) The role of motorcyclist and other driver behaviour in two types of serious accident in the UK, **Accident Analysis and Prevention**, **39**(5), 974-981.

Darby, P., Ison, S. and Raeside, R. (2010) Evaluation of fleet safety interventions. **Proceedings** of the 42nd Universities Transport Study Group conference, Plymouth, UK.

Darby, P., Murray, W. and Raeside, R. (2009) Applying online fleet driver assessment to help identify, target and reduce occupational road safety risks. **Safety Science**, **47**(3), 436-442.

Depaire, B., Wets, G. and Vanhoof, K. (2008) Traffic accident segmentation by means of latent class clustering, **Accident Analysis and Prevention**, **40**(4), 1257-1266.

Doherty, S. T., Andrey, J. C. and MacGregor, C. (1998) The situational risks of young drivers: The influence of passengers, time of day and day of week on accident rates, **Accident Analysis and Prevention**, **30**(1), 45-52.

Flahaut, B., Mouchart, M., Martin, E. S. and Thomas, I. (2003) The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach, **Accident Analysis and Prevention**, **35**(6), 991-1004.

Fraley, C. and Raftery, A. E. (2002) Model-based clustering, discriminant analysis, and density estimation, **Journal of the American Statistical Association**, **97**(458), 611-631.

Häkkänen, H. and Summala, H. (2001) Fatal traffic accidents among trailer truck drivers and accident causes as viewed by other truck drivers, **Accident Analysis and Prevention**, **33**(2), 187-196.

Johnson, R. A. and Wichern, D. W. (2002) *Applied multivariate statistical analysis (5th ed.),* Prentice Hall, Upper Saddle River

Levine, N., Kim, K. E. and Nitz, L. H. (1995) Spatial-Analysis of Honolulu Motor-Vehicle Crashes .1. Spatial Patterns, **Accident Analysis and Prevention**, **27**(5), 663-674.

Li, G., Braver, E. R. and Chen, L. H. (2003) Fragility versus excessive crash involvement as determinants of high death rates per vehicle-mile of travel among older drivers, **Accident Analysis and Prevention**, **35**(2), 227-235.

Linzer, D. A. and Lewis, J. (2007) poLCA: Polytomous Variable Latent Class Analysis. R package version 1.1. Retrieved 20 Feb 2010, from http://userwww.service.emory.edu/~dlinzer/poLCA.

McCall, B. P. and Horwitz, I. B. (2005) Occupational vehicular accident claims: A workers' compensation analysis of Oregon truck drivers 1990-1997. **Accident Analysis and Prevention**, **37**(4), 767-774.

Murray, W., Ison, S., Gallemore, P. and Nijjar, H. S. (2009) Effective occupational road safety programs a case study of Wolseley, **Transportation Research Record**, **2096**, 55-64.

Ng, K., Hung, W. and Wong, W. (2002) An algorithm for assessing the risk of traffic accident, **Journal of Safety Research**, **33**(3), 387-410.

Quddus, M. A., Noland, R. B. and Chin, H. C. (2002) An analysis of motorcycle injury and vehicle damage severity using ordered probit models, **Journal of Safety Research**, **33**(4), 445-462.

Savolainen, P. and Mannering, F. (2007) Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes, **Accident Analysis and Prevention**, **39**(5), 955-963.

Skyving, M., Berg, H. Y. and Laflamme, L. (2009) A pattern analysis of traffic crashes fatal to older drivers, **Accident Analysis and Prevention**, **41**(2), 253-258.

Smith, W. R. (1956) Product Differentiation and Market Segmentation as Alternative Marketing Strategies, **Journal of Marketing**, **21**(1), 3-8.

Statistical Innovations (2008) [online]. [Accessed 24 Feb 2010]. Available from World Wide Web : http://www.statisticalinnovations.com

Stone, M. and Broughton, J. (2003) Getting off your bike: cycling accidents in Great Britain in 1990-1999, **Accident Analysis and Prevention**, **35**(4), 549-556.

Ulleberg, P. (2001) Personality subtypes of young drivers. Relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign, **Transportation Research Part F: Traffic Psychology and Behaviour**, **4**(4), 279-297.

Wallington D. (2009) Occupational Road Risk Management - A Competitive Advantage? **Presented** at the *NIOSH Global Road Safety for Workers Conference*, Washington DC, 16-18 February 2009,

Wedel, M. and Kamakura, W. A. (1998) *Market Segmentation: Conceptual and Methodological Foundations,* Klewer, Boston.

Xu, R. and Wunsch, D. (2005), Survey of clustering algorithms, **IEEE Transactions on Neural Networks**, **16**(3), 645-678.