NZ Society of Actuaries Conference
October 2018

An actuarial approach to modelling vehicle injuries and fatalities

Jeremy Holmes, Marcella Lau and Aaron Park

MELVILLE JESSUP WEAVER
Willis Towers Watson Alliance Partner

Abstract

Traditional modelling of vehicle injuries and fatalities has focused either on large scale time series analysis e.g. attempting to assess macro economic impacts on the road toll, or detailed technical modelling applied to specific issues e.g. roundabout safety. There is general consensus that the improved quality of vehicles and roads is a major contributor to the falling road toll over the last few decades. However, recent years have shown a slight reversal of this trend in New Zealand.

In this paper we attempt to understand the movements in the road toll by applying the actuarial approach. That is, by defining exposure measures and using multivariate analysis to understand the impacts on accident frequency and severity – much in the same way as motor insurance is priced. There is a particular focus vehicle safety features and distance travelled.

Foreword

The purpose of this paper is not to provide an exhaustive analysis of the causes of vehicle injuries and fatalities. The data available for this analysis was voluminous but limited in detail - Appendix A summarises the data that was provided. The focus here is very much on the safety information for each vehicle in New Zealand and how that corresponds to the safety information for vehicles involved in fatal accidents or serious injuries. There are many important factors which have not been considered, such as:

- The characteristics of the driver(s)
- The location of each registered vehicle
- The location and details of each accident e.g. weather/road conditions, influence of speed, alcohol or drugs.

Additionally, a large number of material approximations and assumptions have been made where the nature of the data has required this. For these reasons, the results in this report should be considered indicative rather than definitive. In Appendix C we discuss a number of improvements which could be made to the modelling, and the data which would be required to facilitate this.

The purpose of this paper is not to definitively quantify the impact on the road toll of vehicle safety features, but rather to demonstrate the actuarial method and its applicability to an analysis of the road toll. Our hope is that this paper will generate an interest amongst actuaries in this field and that ultimately more detailed data will be made available to enable more rigorous and comprehensive actuarial analyses to be undertaken.

Comparing to other published information

Many of the statistics given in this paper have been published in other forums as well. For example, the total number of kilometres travelled by NZ vehicles is a figure published regularly by the NZ Transport Agency. However, the bases upon which these figures have been determined can vary from one source to the next. This is particularly the case with odometer readings where a significant amount of estimation and approximation is required. The analyses in this paper are based on data provided by NZTA, but NZTA has not been involved in the data cleaning or analysis.

Contents

Appendices

1 Summary

1.1 Purpose and scope

The purpose of this paper is to test the value of the actuarial method to the analysis of vehicle injuries and fatalities. The analysis here is undertaken in the context of an increasing road toll in recent years. Whilst the recently increasing number of road fatalities is particularly concerning, the focus of this paper is not so much on understanding the cause of that single increase but on generally understanding the (positive and negative) factors of that affect the number of injuries and fatalities, and particularly as they pertain to the vehicle.

We were supplied with crash data going back to around 1980. However, this paper focuses on using VKT as an exposure measure, for which we only have reliable data back to around 1997. Therefore, the analysis here is focused on the period from 1997 to 2017 and focuses only on injuries and fatalities to the occupants of the vehicle(s) i.e. excludes pedestrian or cyclist fatalities.

1.2 Methodology

In order to understand the causes of vehicle injuries and fatalities we have broken the analysis into two components:

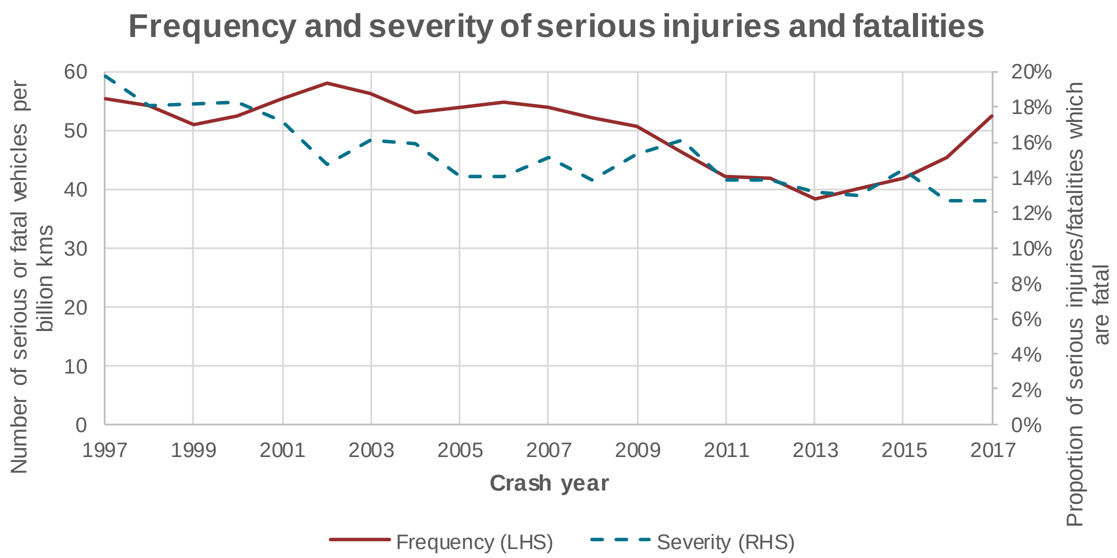
- Frequency: the number of vehicles in which serious injuries or fatalities have occurred, expressed per distance travelled.
- Severity: the proportion of serious accidents which result in a fatality.

In Section 2 we discuss the methodology in more detail.

1.3 Key results

1.3.1 20 year trend

The chart below shows the observed frequency and severity of serious injuries and fatalities from 1997 to 2017.



The frequency of serious accidents (when expressed per kilometre travelled) generally decreased from the early 2000s to around 2013, after which there is a sharp increase. The severity (i.e. proportion fatal) has generally decreased over the entire period, with no obvious recent increase as is observed for the frequency.

In the sections below we discuss some of the factors affecting frequency and severity.

1.3.2 Frequency

The key factors identified as affecting the frequency of serious accidents are:

- The Used Car Safety Rating (UCSR) or ANCAP rating of the vehicle
- The year of manufacture and age of the vehicle (at the time of the accident)
- The mass and vehicle class (e.g. motorcycle, passenger vehicle, truck)

The general downward trend in frequency over the period is largely explained by the increased proportion of better rated vehicles being used on our roads. In Section 6.1.2 we have attempted to quantify the impact on frequency of driving better rated vehicles.

Where there is no UCSR or ANCAP rating attached to a vehicle we have attempted to model the frequency using a combination of the vehicle year and age. The relationship with vehicle age is as expected i.e. older vehicles are involved in a greater number of serious accidents (per kilometre) than newer vehicles – this may be related to changes in the driver as a vehicle ages (e.g. younger drivers in older vehicles). The relationship with vehicle manufacture year however is more complex. In Section 6.1.4 we identify vehicles of particular eras which appear to be have higher frequencies.

An important factor is the frequency of motorcycle accidents The table below considers the frequency of serious motorcycle accidents by manufacture year and age of the motorcycle at the time of the accident. Each diagonal represents a year of observation e.g. the latest diagonal shows accident frequencies during 2017.

Fatal or serious vehicles per billion kms - adjusted for vehicle mass

Vehicle	Vehicle age (years)																									
year:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1990								532	663	537	603	748	1,017	1,928	859	1,004	290	641	379	360	867	428	-21	464	580	674
1991							436	841	264	1,607	262	261	302	1,580	290	-21	357	773	428	498	573	-22	-22	-21	1,382	811
1992						404	456	853	1,230	466	785	498	2,247	504	501	-22	468	411	551	-21	-22	733	688	1,921		
1993					1,300	602	511	949	-22	-22	-21	-22	1,802	564	-22	478	-22	519	652	836	-21	-21		-21		
1994				-20	871	-21	892	476	-21	553	567	-21	398	-21	-21	518	962	-22	664	-21	-21	547	758	726		
1995			569	-20	-20	600	-20	251	-20	751	682	306	-21	-21	710	-21	340	-21	-21	560	467	1,899	991			
1996		509	-21	537	800	214	237	1,143	-20	276	-21	-21	-21	311	1,040	352	-21	381	-21	892	1,490	3,591				
1997	838	527	-21	389	809	232	520	745	462	1,587	515	274	294	1,509	-21	1,951	361	913	1,028	2,330	3,830					
1998		1,074	194	426	467	-21	244	-21	-21	510	540	226	312	363	-21	367	-21	486	2,240	1,061						
1999	704	325	381	644	470	1,005	477	468	949	225	-21	240	294	-21	-21	-21	384	411	2,363							
2000	587	167	208	720	1,330	244	-21	738	224	-21	1,722	-21	304	309	-21	725	1,624	974								
2001	276	140	159	742	357	-21	159	340	585	401	681	462	218	258	553	1,562	2,521									
2002	241	548	443	140	141	-21	316	329	751	-21	-21	-21	232	236	813	609										
2003	465	1,198	258	392	263	268	449	675	365	385	193	681	435	425	1,477											
2004	750	841	-21	294	89	828	111	279	465	145	144	-21	769	1,699												
2005	592	370	476	428	583	524	324	441	345	607	385	817	1,940													
2006	274	67	71	697	94	102	758	251	270	592	1,650	838														
2007	449	474	29	267	235	115	418	135	311	1,239	1,527															
2008	625	480	151	364	397	127	298	236	852	1,921																
2009	225	341	571	73	379	523	322	455	1,131																	
2010	645	-21	411	218	359	526	874	1,003																		
2011	196	202	100	110	260	566	467																			
2012	404	-22	227	110	550	1,089																				
2013	182	77	196	449	1,070																					
2014	135	129	866	729																						
2015	234	92	1,091																							
2016	399	277																								
2017	178																									

In recent years there has been a strong increase in the exposure (i.e. kilometres travelled) to motorcycle use for new (2015/16+) motorcycles. This partly explains the increase in motorcycle accidents, although the frequency (i.e. per kilometre) for very new motorcycles doesn't stand out as unusually high. There has however, been a strong increase in motorcycle accident frequencies for older motorcycles. Again this may be related to the age or other characteristics of the driver.

1.3.3 Severity

The factors identified as affecting the severity of accidents are:

- The UCSR ratings for the primary vehicle
- The vehicle year for the primary vehicle (if there is no UCSR rating)
- The mass and type of the primary vehicle
- The mass and type of the other vehicle(s) involved

In Sections 6.2.3 and 6.2.4 we explore the impact of improved UCSR ratings and later models over time as predictors of severity. The table below shows what our modelling has identified as the important factors for severity as they relate to the other vehicle(s) involved.

Multiple vehicle accidents in which the primary vehicle does not have a UCSR
Relative risk identified by the GLM for severity

		Accidents involving other:					
Mass of the heaviest other vehicle	Vehicle mass ratio	Motorcycles	Cars	Vans	SUVs	Trucks	Buses
0-1,000kg	0-75%	164%	-	149%	-	-	-
1,000-1,500kg	0-75%	214%	130%	194%	192%	367%	234%
1,500-2,000kg	0-75%	254%	155%	231%	229%	437%	278%
2,000-3,000kg	0-75%	270%	164%	245%	243%	464%	296%
3,000-10,000kg	0-75%	192%	117%	174%	173%	330%	210%
10,000kg+	0-75%	211%	128%	191%	-	362%	230%
0-1000	75%-500%	229%	-	-	-	-	-
1000-1500	75%-500%	-	182%	271%	268%	512%	326%
1500-2000	75%-500%	355%	216%	322%	319%	609%	388%
2000-3000	75%-500%	377%	229%	342%	339%	647%	412%
3000-10000	75%-500%	-	163%	243%	241%	459%	293%
10000+	75%-500%	-	179%	-	-	504%	-
0-1000	500%+	-	-	-	-	-	-
1000-1500	500%+	-	231%	-	-	-	-
1500-2000	500%+	450%	274%	409%	405%	-	-
2000-3000	500%+	478%	291%	434%	430%	-	-
3000-10000	500%+	-	207%	308%	305%	583%	372%
10000+	500%+	-	227%	338%	335%	640%	408%

The table shows:

- Accidents involving larger vehicles (buses and particularly trucks) are more likely to result in a fatality for the primary vehicle occupants.
- As the mass of the other vehicle increases (whatever that vehicle is) the likelihood of fatality increases.
- Not only the overall mass but also ratio of the masses of the vehicles is important.

Whilst these findings aren't unexpected, it is interesting to attempt to quantify just how important the other vehicle(s) are to the likelihood of surviving a serious accident. This result supports NZTA's focus on encouraging vehicle purchasers to choose the 'safer pick' vehicles which are both better for the vehicle occupants as well as the other parties involved.

In Section 6.2.5 we further explore the implications of this table.

1.4 Reliances and limitations

This paper is intended to generate discussion around the actuarial method for vehicle injuries and fatalities and to suggest areas for further research. This paper is not actuarial advice upon which to base decisions. We advise against drawing any conclusions from the results in this paper. This is particularly the case given that we have focused entirely on the characteristics of the vehicle in the absence of any information in regard to other factors such as the driver or the causes of the accident.

This paper may be reproduced and distributed in its entirety, so long as Melville Jessup Weaver is acknowledged as the source. However, Melville Jessup Weaver will not be held responsible if any party chooses to rely on the results in this paper.

2 The actuarial method

2.1 Insurance pricing

The actuarial method can mean different things in different contexts. In this paper when we refer to the actuarial method we mean the approach used by actuaries to risk rate insurance premiums. For private motor insurance this might involve:

- Defining an exposure measure e.g. the time exposed to risk (duration of the insurance policy)
- Defining an observation i.e. a claim
- Hypothecating a list of variables which might influence the likelihood or cost of a claim e.g. driver age/sex, vehicle age/make/model, sum insured
- Calculating the exposure for each combination of variables.
- Summing the number of claims for each combination of variables
- Calculating the claim frequency for each combination of variables i.e. observed number of claims divided by exposure
- Calculating the claim severity (i.e. average claim size) for each combination of variables
- Employing a multivariate analysis (often a generalised linear model) to understand the impact of each factor or combination of factors on claim frequency and/or severity.

This method enables the actuary to determine a 'risk price' for each combination of variables. For example, the expected cost of claims for a 25 year old male driving a 2005 Subaru Legacy worth $15,000 and housed in a garage in Henderson would be the estimated frequency multiplied by the estimated severity for that particular combination of risk factors.

2.2 Exposure measures

The exposure measure used to risk rate motor insurance has, until recently, been almost exclusively the policy duration in days/years. Developments in telematics devices and smartphones have enabled the use of mileage as an exposure base or rating factor in some overseas markets, although telematics developments in New Zealand have been limited.

Fortunately for the analysis in this paper the NZ Transport Agency holds a wealth of data on mileage readings for almost all vehicles registered in the country. This data is collected at warrant of fitness checks by vehicle inspectors and is of varying quality – see Appendix B. Nevertheless, it is a useful source of data and, because the data is collected and stored at an individual vehicle level, enables us to consider using kilometres travelled as an exposure base i.e. to analyse accident frequencies per kilometre travelled.

2.3 Frequency and severity

In analysing movements in the road toll there isn't really a concept of severity like there is for insurance claims – there are not varying degrees of fatality. However, there is a concept of injury severity. The accident data we have used for this analysis has been extracted from the Casualty Analysis System (CAS) and is discussed in Appendix A. The data contains summaries of all the minor/major injuries and fatalities on NZ roads. As an analogy to insurance claim severity we have considered the proportion of serious injuries which result in a fatality.

Effectively, we have adapted the insurance pricing approach to assess:

- The frequency of serious accidents per kilometre travelled
- The proportion of serious accidents resulting in a fatality
- The factors (specifically around vehicle safety) affecting the two points above.

2.4 Granularity

A fundamental element of the actuarial method for insurance pricing is that all risk factors need to be analysed concurrently. For example, it is of little use to know that 40% of policyholders are male and 10% of policyholders drive Subarus, without knowing specifically which males drive Subarus. More generally, in order to undertake a multivariate analysis it is necessary to break the exposure and claims observations down into subsets for every possible combination of risk factors.

The same is true when applying the actuarial method to an analysis of the road toll. Whilst numerous studies have considered the impact of overall Vehicle Kilometres Travelled ('VKT') on fatality numbers (see Section 3) the actuarial method differs in that it seeks to split VKT according to every possible combination of risk factors and employ an exposure based multivariate frequency/severity approach to predict the numbers of injuries and fatalities and the causes thereof.

3 Previous investigations

There have been a number of quantitative studies in recent years to investigate trends in the NZ road toll. Two notable studies are:

- *Econometric Analysis of the Downward Trend in Road Fatalities since 1990* (July 2013) by Infometrics
- *Qualitative and Quantitative Analysis of the New Zealand Road Toll* (March 2017) by Deloitte Access Economics

The first of these two studies was commissioned to help to understand the general downward trend in road fatalities, whilst the second was commissioned to help understand the recent increase.

Also, a number of reports have been released by Monash University with a general focus on crashworthiness (the Used Car Safety Ratings are published by Monash University).

3.1 Infometrics study

The Infometrics study considered the impact of vehicle safety via a time series analysis. A measure was defined which represented a number of factors (including vehicle safety) and this measure was used to help explain the annual number of vehicle fatalities and/or injuries. Vehicle kilometres travelled was considered as an explanatory factor, but only at an aggregate level. For example, VKT was not analysed separately by vehicle year and crashworthiness rating.

The Infometrics study attributed a portion of the reduction in the road toll to overall vehicle improvements (improved crashworthiness and fewer motorcycles). However, the study did not quantify the specific changes in crashworthiness (e.g. moving from a 4 to 5 star crashworthiness rating) and how these have impacted the road toll.

3.2 Deloitte study

The Deloitte study also considered VKT as a predictive factor for the number of crashes and indeed identified it as significant. The number of motorcycle registrations was also identified as significant. However, like the Infometrics study, VKT was used only at an aggregate level. The Deloitte study did not analyse VKT separately for vehicles of different crashworthiness or for separate vehicle types (e.g. VKT for trucks vs. motorcycles).

The Deloitte study noted that a more detailed breakdown of VKT would be beneficial for future analyses.

3.3 Monash studies

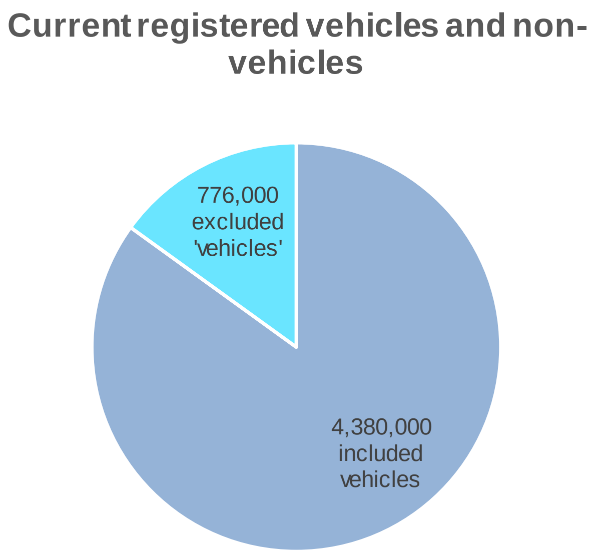
Monash University has undertaken a considerable amount of research into the crashworthiness of vehicles in Australia and New Zealand. A lot of this research is based upon Monash's analysis of a very large database of crash records in Australia and NZ. The focus of this research is largely on the outcome for people involved given that there has been a crash, rather than the probability that there will be a crash in the first place.

In order to address the question of the probability that there will be a crash it is necessary to understand not just the number of accidents but also how much travel has been undertaken without accident (as this paper intends to address).

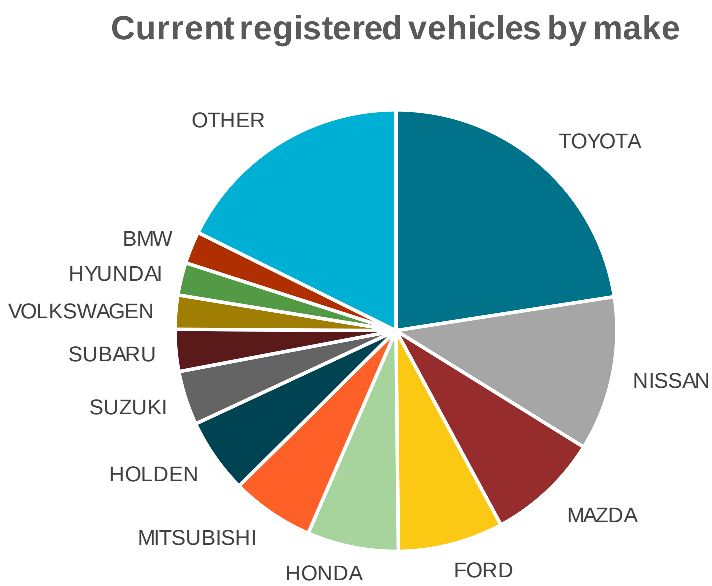
4 The New Zealand vehicle fleet

4.1 Current registered vehicles

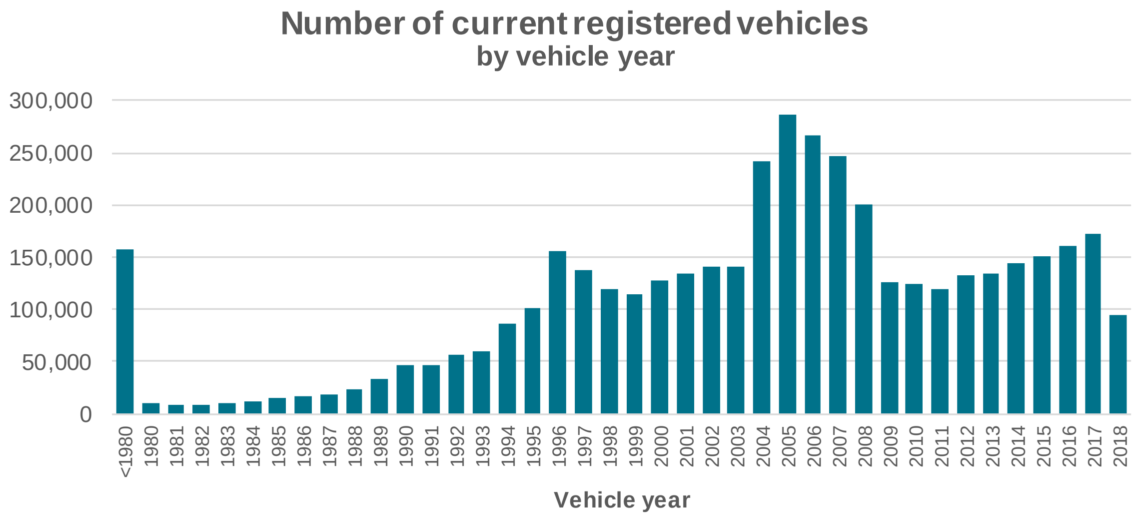
The NZ vehicle fleet comprises a little over 5 million currently registered vehicles. However, many of these are not relevant to our analysis in this paper – there are a significant number of trailers registered, as well as numerous farm vehicles which spend little time on public roads. The chart below shows the portion of current vehicles which are excluded from this analysis. Unless otherwise stated, all results in this paper refer to the 'included' vehicles only.



Drilling down into the included vehicles we see that the bulk of vehicles in NZ are mainstream Japanese brands. European branded vehicles comprise a relatively small proportion of the total fleet, although that proportion is changing over time (see Section 4.2.1).

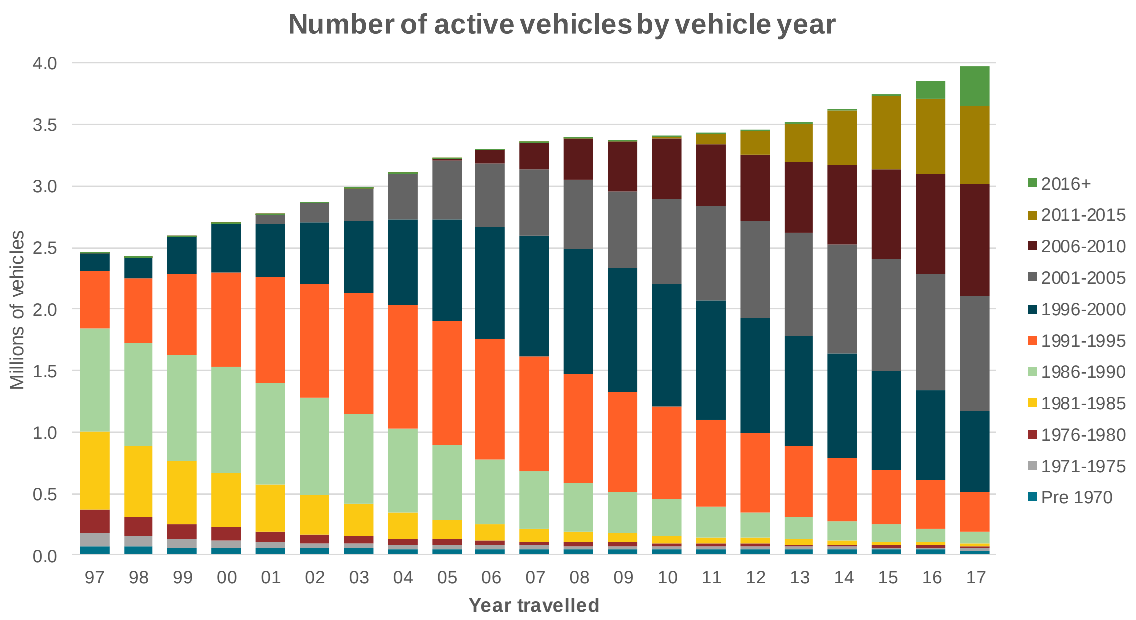


NZ has a relatively old vehicle fleet by global standards. Breaking down our current fleet by year we see that NZ has a significant number of older vehicles – in particular there are more than 150,000 pre-1980 vehicles still registered to be used on NZ roads (although the number with a current warrant of fitness may be less).

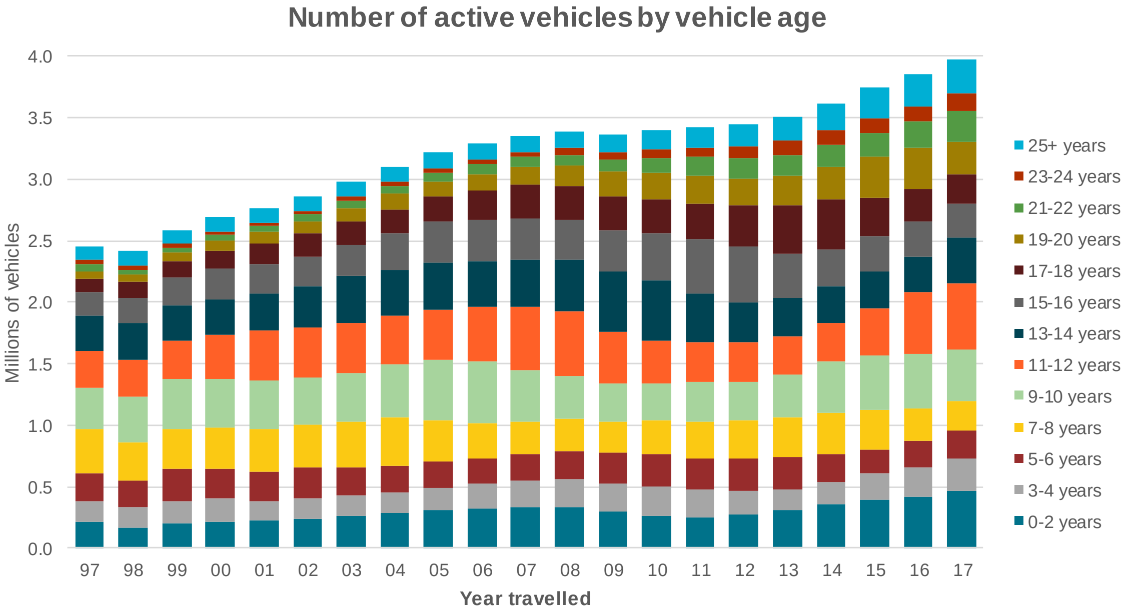


4.2 Changes over time

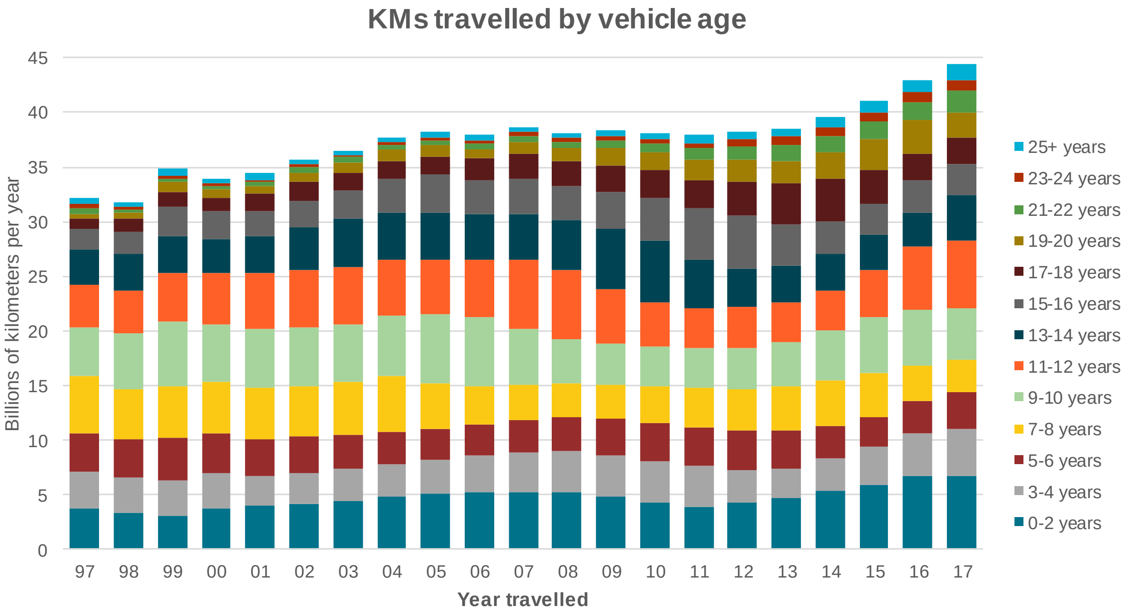
By analysing mileage readings for all currently registered and deregistered vehicles we were able to see how the NZ vehicle fleet has changed over time. The chart below looks at the number of active vehicles by vehicle year. We have defined active vehicles as those which recorded a mileage reading during the year (including where we have estimated readings – see Appendix B).



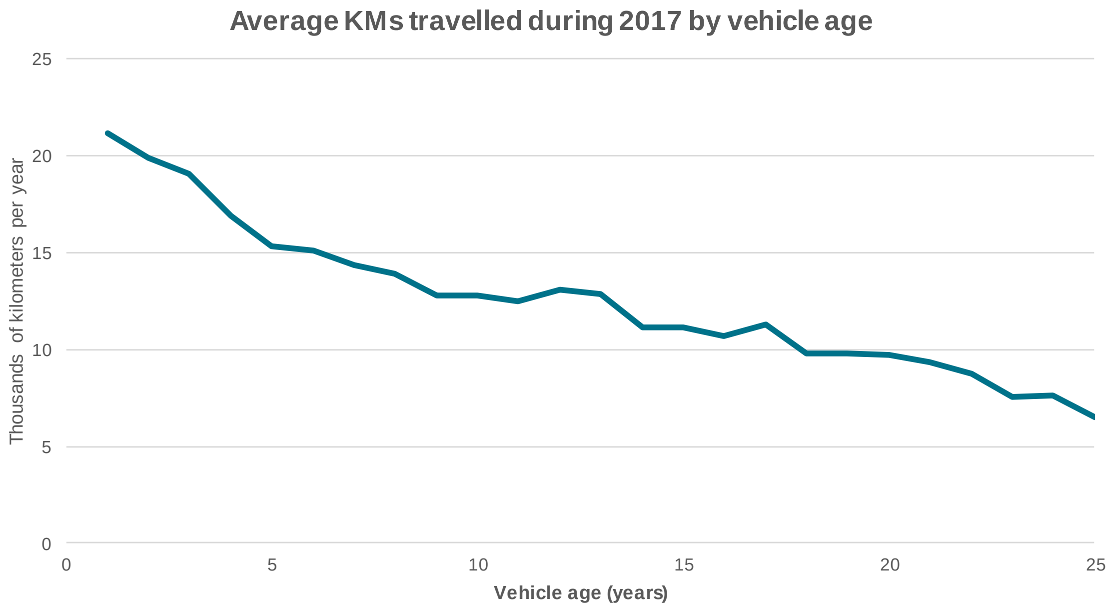
Slicing the data instead by vehicle age we see that there are a growing number of older vehicles being used (although there are increasing numbers of vehicles in other age bands as well).



When we look at the number of kilometres travelled by these vehicles we see that older vehicles represent a relatively smaller portion of kilometres travelled.

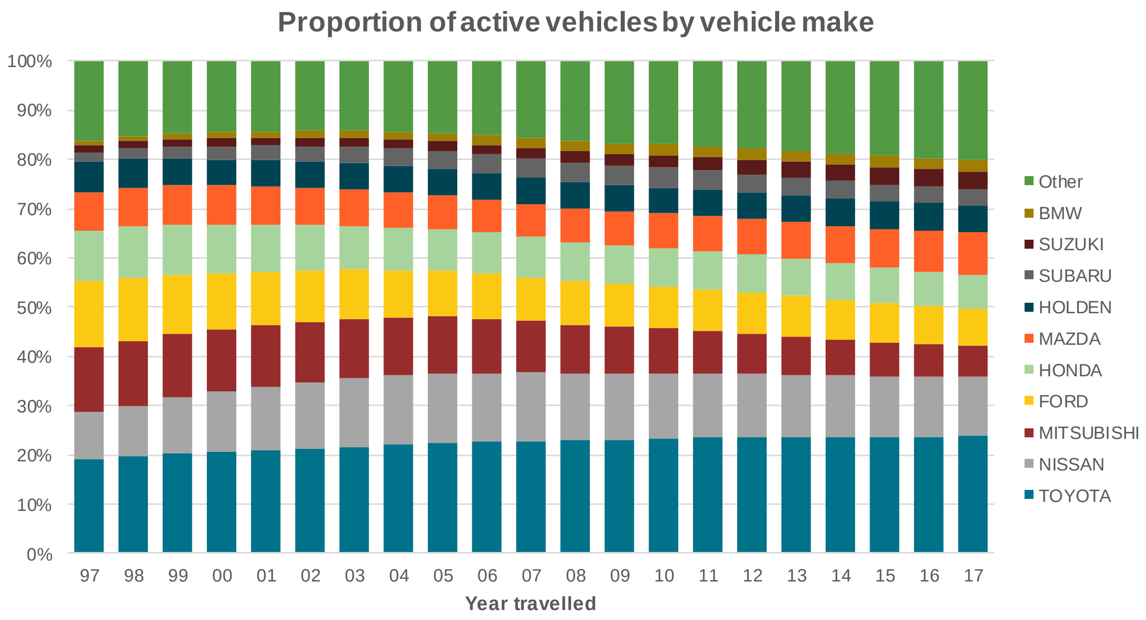


Bringing the two charts above together, and filtering on 2017 as an example, we see that older vehicles generally cover fewer kilometres. This is a widely known effect and is likely due to the changing nature of the driver over the vehicle's lifetime e.g. the vehicle may start as part of a company fleet for the first few years and later become an occasional vehicle for a young driver. It is important to consider this for any analysis of the impact of safety features. Newly introduced safety features will have a proportionately higher representation per kilometre than per vehicle. Indeed this is a key reason for our analysis of kilometres travelled as an exposure measure.

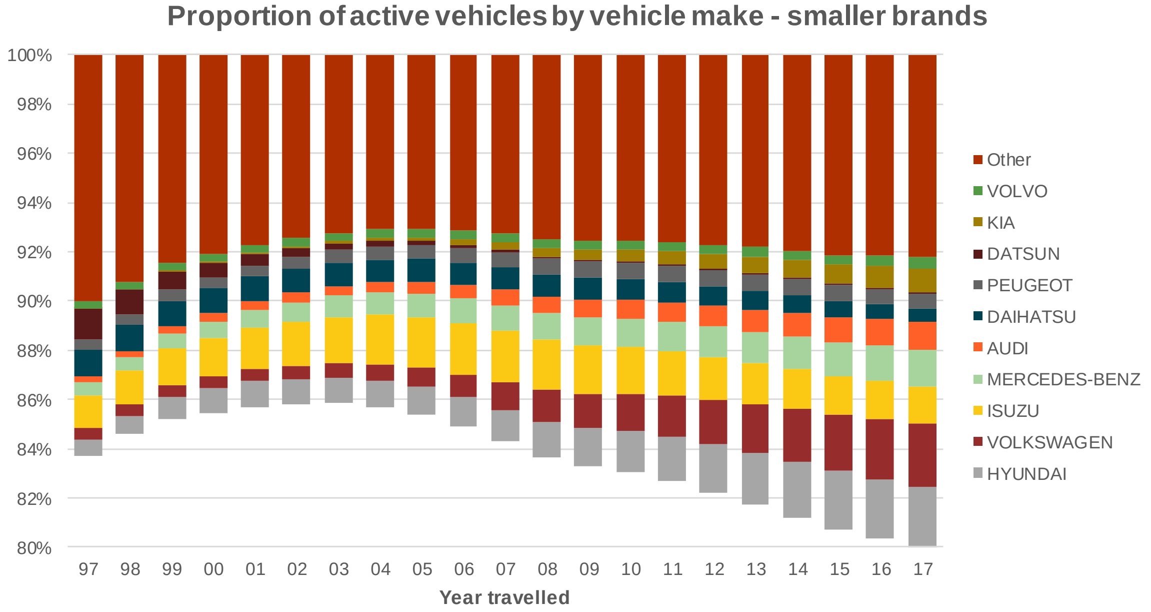


4.2.1 *Vehicle make changes over time*

The chart below looks at how vehicle manufacturers have changed over time. With some exceptions, the large Japanese brands have ceded market share to European and other smaller brands. Considering the period from 1997 to 2017 as a whole, BMW is the only European brand in the top 10.



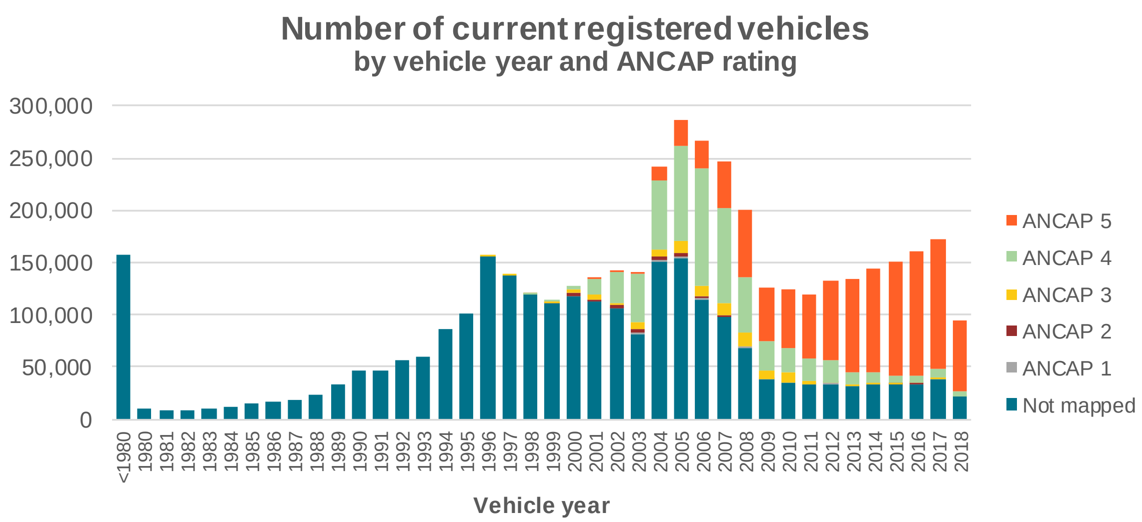
Zooming in on the 'Other' category we see that these smaller brands have grown in market share.



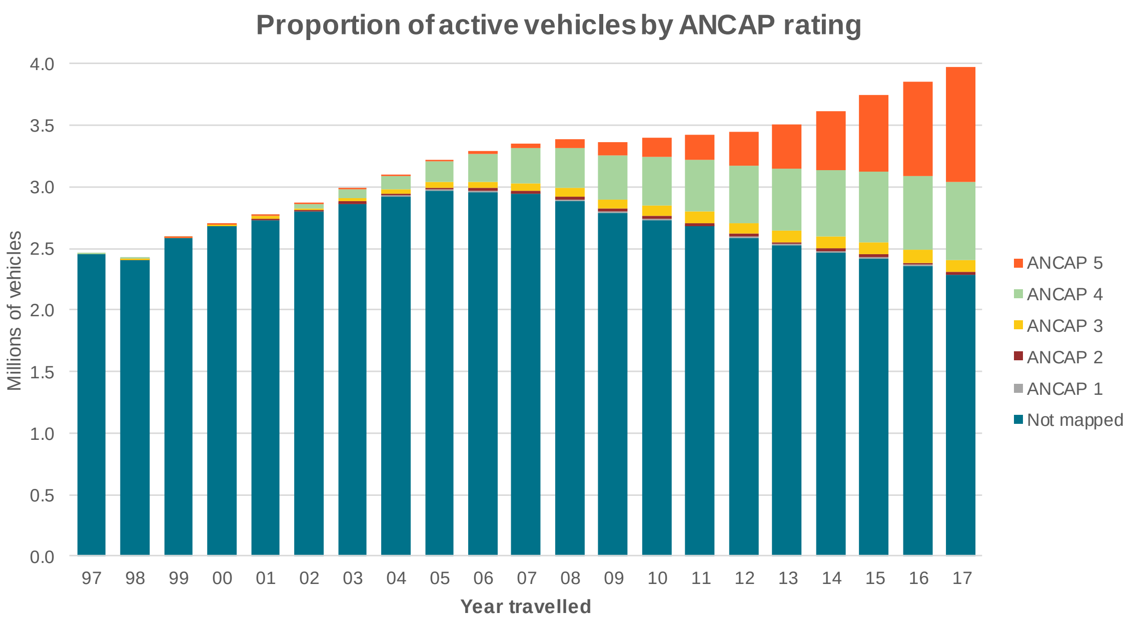
Whilst these summaries of vehicle make are not directly applicable to the analysis in this paper (we haven't considered vehicle make as a potential explanatory factor) they do provide some useful context to changes in the NZ fleet over time.

4.3 ANCAP ratings

The chart below looks at the distribution of NZ's vehicle fleet by ANCAP rating. For the majority of vehicles manufactured in the last decade or so we have been able to map an ANCAP rating to the vehicle. Going back further than this the number of ANCAP mapped vehicles is fewer – not surprising given that the focus of ANCAP ratings is on new vehicles. It is important to note that the ANCAP testing methods (and resulting star ratings) change over time – a 5 star 2017 vehicle is not the same as a 5 star 2007 vehicle. Whilst this effect is not illustrated in this chart it has been considered as part of the analysis described later in this paper.

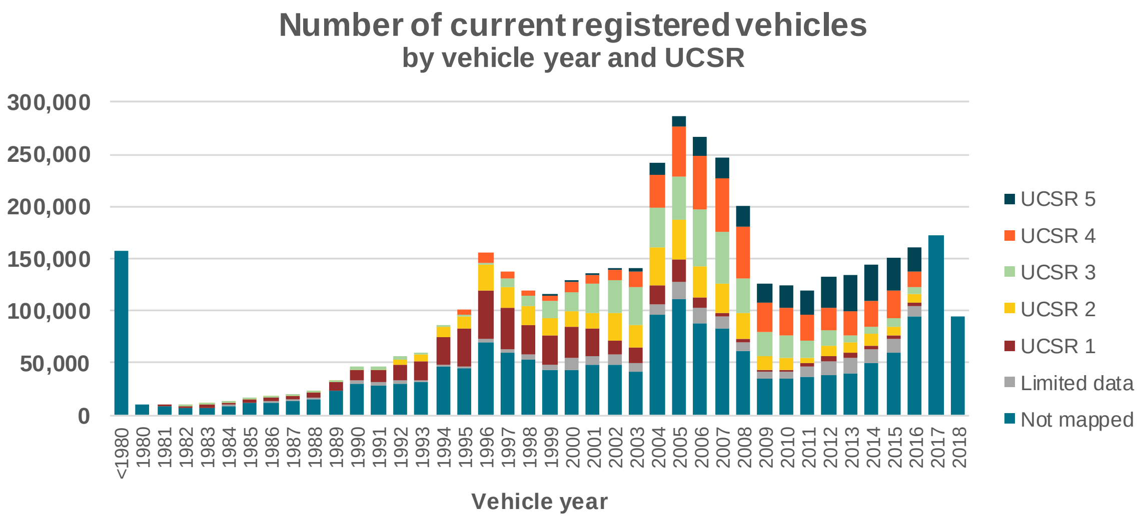


Looking at the proportions by ANCAP rating of active vehicles over time we see that a relatively small portion of the vehicles on NZ roads have an ANCAP rating – a function of our relatively old vehicle fleet – although for vehicles active more recently this proportion is growing.



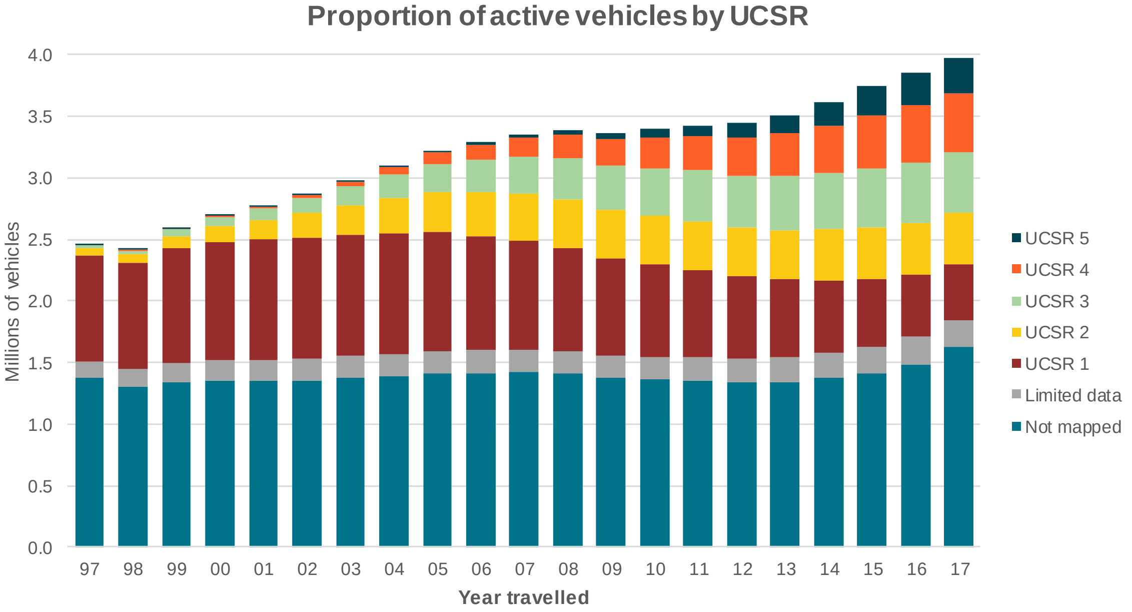
4.4 UCSR ratings

Whilst ANCAP ratings focus largely on new vehicles, Used Car Safety Ratings have a much stronger focus on the performance of older cars. The chart below breaks down our current vehicle fleet by UCSR crashworthiness rating. Few new vehicles have UCSRs (as one might expect) but UCSRs do cover a much larger portion of older vehicles. The trend towards higher ratings for (relatively) newer vehicles is evident.



In order for a UCSR crashworthiness star rating to be allocated to a vehicle model, a minimum sample size is required. The 'limited data' category above is where a score has been estimated for that vehicle but the sample size is too small to allocate an overall crashworthiness star rating.

The chart below considers changes in the vehicle fleet over time by UCSR.



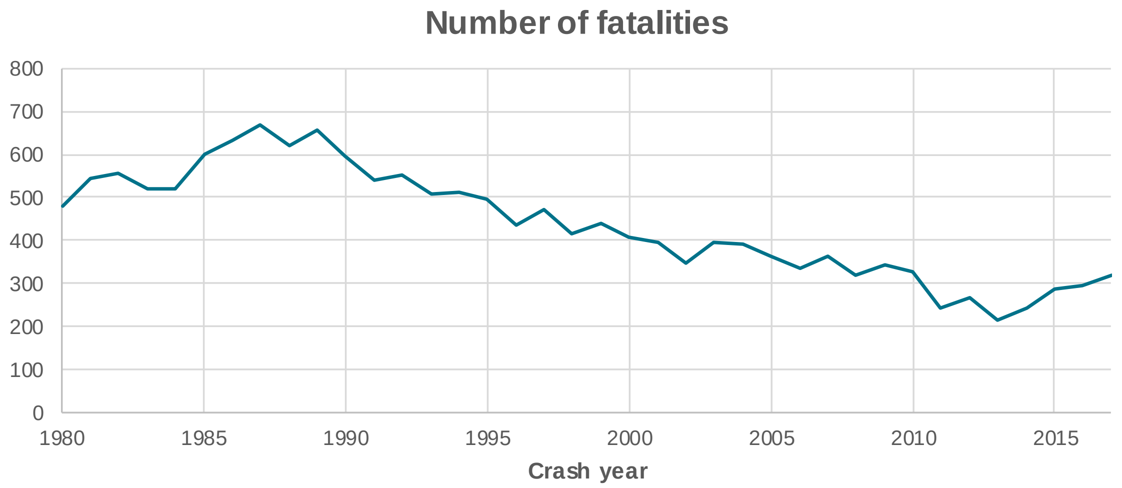
The generally improving UCSR figures for our vehicle fleet is clear.

Interestingly, where the majority of ANCAP rated vehicles receive either a 4 or 5 star ANCAP rating, the UCSR figures show a greater spread of star ratings from 1 to 5. Whilst there is little useful comparison between, say, 5 star UCSR vehicles and 5 star ANCAP vehicles (the scoring systems are fundamentally different), the greater breadth of UCSRs may prove to be a better predictor of fatality/injury rates. In any case, the analysis in Section 6 focuses more on the actual UCSR scores than the particular star rating to which a score has been categorised.

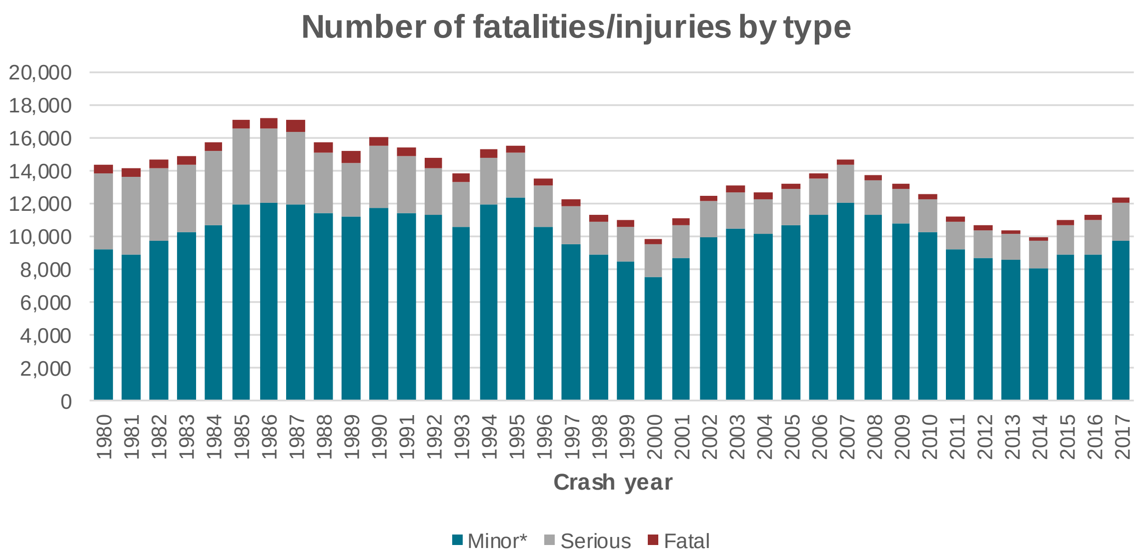
5 New Zealand accident history

5.1 Fatalities and injuries

The chart below shows the number of fatalities on NZ roads since 1980. After increasing to a peak in the late 80s fatality numbers decreased steadily for a few decades until around 2013, after which fatalities increased, albeit slightly, year on year for the next four years.



Combining fatalities with injury numbers we get the chart below. Injury numbers also peak in around the mid to late 1980s. However, there is a second peak around 2007, after which injury numbers declined until around 2014 and then slightly increased in line with fatality numbers. Unfortunately, there is likely to be considerable underreporting of minor injuries, and the degree of underreporting may vary over time. For this reason, the focus of the analysis in this paper is on fatalities and serious injuries only.

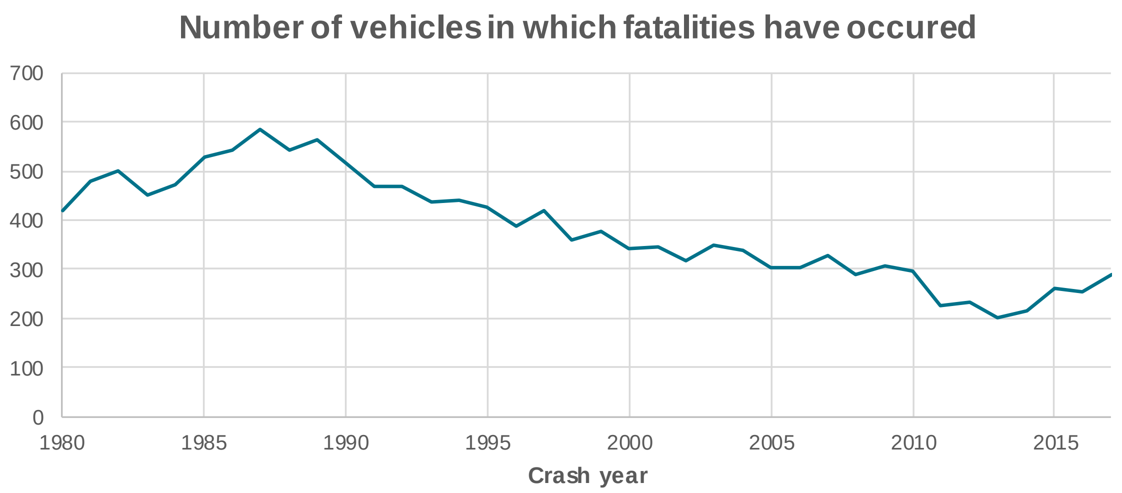


*Reporting of minor injuries is likely to be materially understated

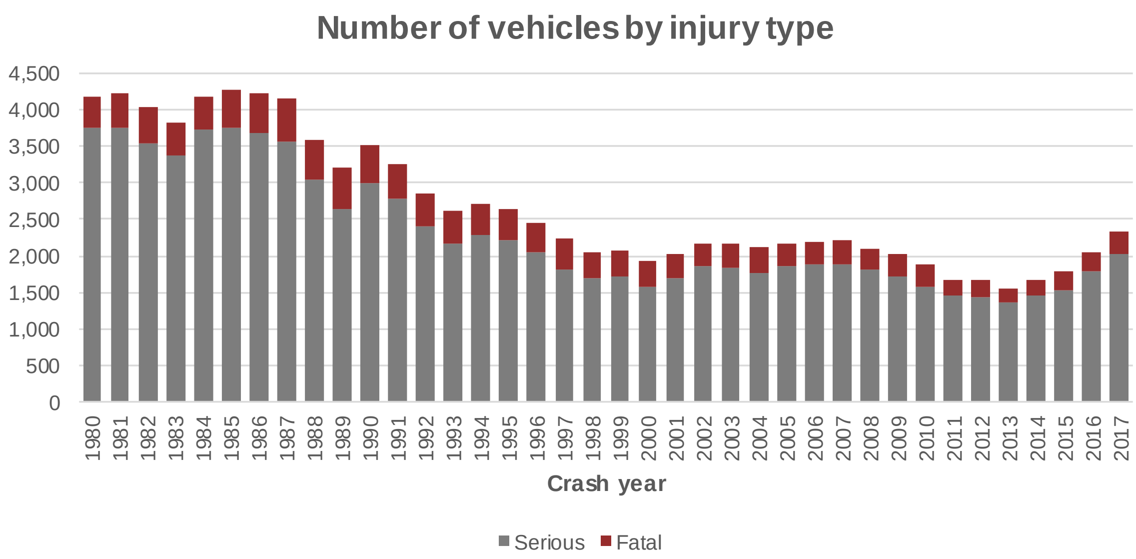
5.2 Fatal/injury vehicle numbers

In some cases there are multiple fatalities or injuries within a single vehicle. It is useful to also consider the number of 'fatal vehicles' i.e. the number of vehicles in which one or more fatalities occurred. Under this measure each fatal accident represents a single observation (unless there were also fatalities in one or more other vehicles, in which case multiple observations are counted).

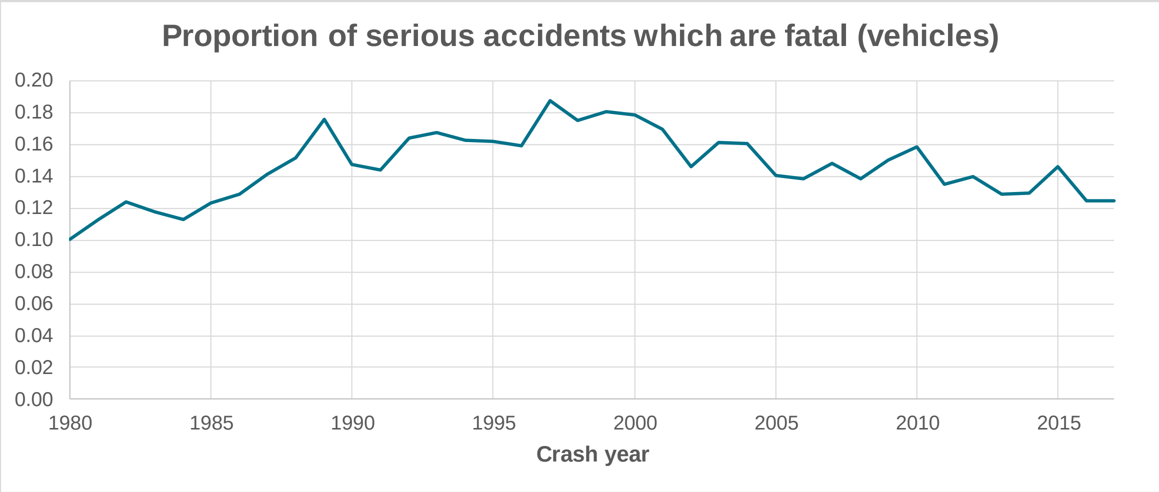
The chart below shows the number of vehicles in which fatalities have occurred. The pattern is very similar to the total number of fatalities.



Similarly we can also look at the number of vehicles in which serious injuries or fatalities have occurred.



It may be beneficial to treat fatalities as a special case of a serious injury i.e. where the injury is severe enough to cause death. This is the frequency/severity concept discussed in Section 2. The chart below looks at the proportion of serious accidents which are fatal (again considering each vehicle as a single observation). In this case there is a gradual decline in the proportion from around the late 1990s to 2017.

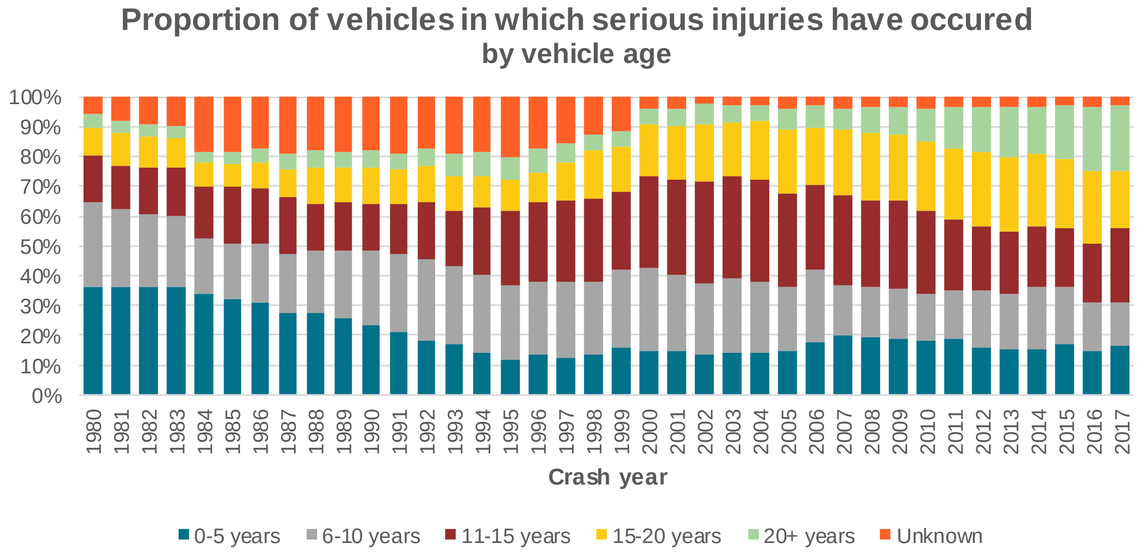


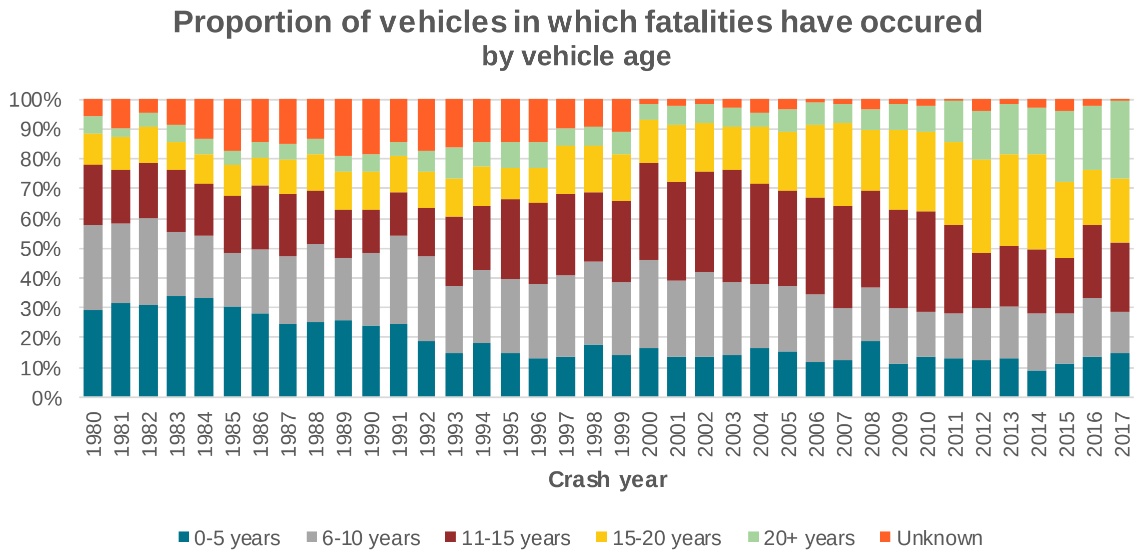
It is not immediately obvious whether one might expect developments in vehicle safety to increase or decrease this ratio. Both the numerator (fatalities) and the denominator (serious injuries + fatalities) will be reduced with safety improvements. In this sense the frequency/severity concept doesn't translate perfectly from the world of insurance pricing to fatality/injury modelling. Nevertheless, there is value in considering whether insights can be drawn by considering these metrics separately.

5.3 Fatalities and injuries by vehicle age

The charts below show the proportions of serious and fatal vehicles broken down by vehicle age at the time of the accident. Recording of vehicle details in CAS has improved over time and the number of unknown vehicles (i.e. where the vehicle year wasn't recorded) has reduced (see Section 5.6 for more on this).

There is perhaps a slightly greater proportion of accidents occurring in older vehicles in more recent years, although there are other impacts which might be affecting the figures presented here. In Section 6 we address this (to some extent) with multivariate analysis.



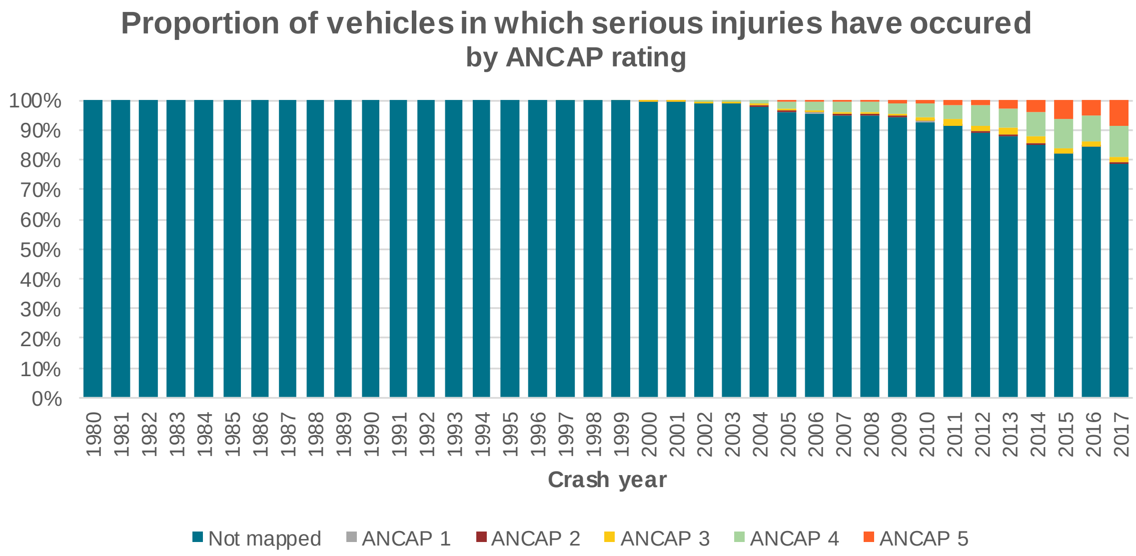


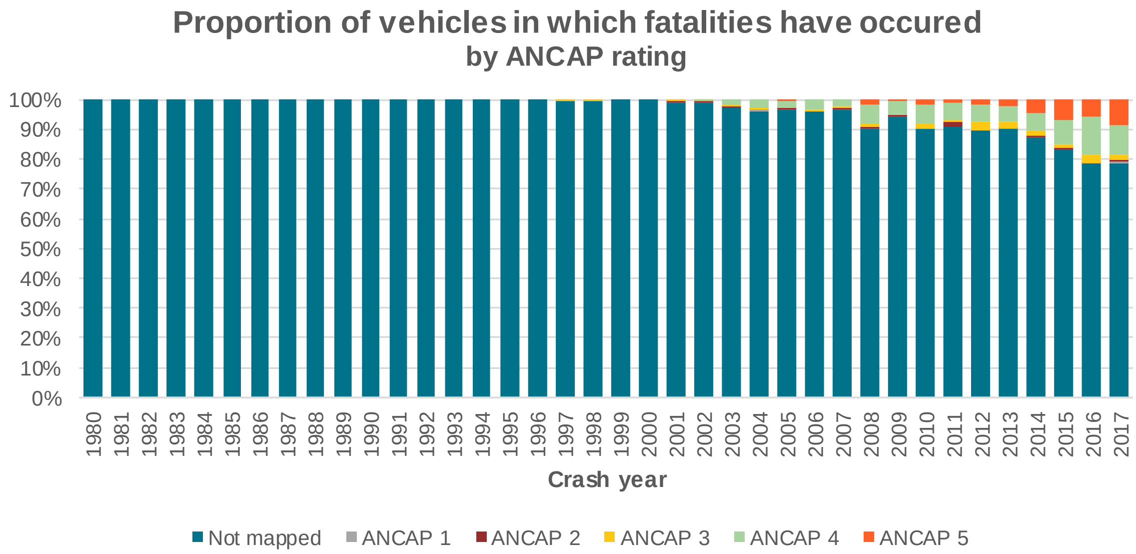
5.4 Fatalities and injuries by ANCAP rating

Here we look at the number of serious injuries and fatalities by ANCAP rating. The first impression from these charts is that ANCAP rated vehicles appear to be underrepresented in accident numbers compared to vehicle numbers or kilometres in Section 4.3. This could be due to:

- Difficulties matching CAS records to ANCAP ratings due to the lower quality of vehicle information in CAS compared to the MVR (in Appendix C we discuss how this could be improved).
- ANCAP rated vehicles improving outcomes for occupants in the event of a crash and/or reducing the likelihood of being involved in a crash.

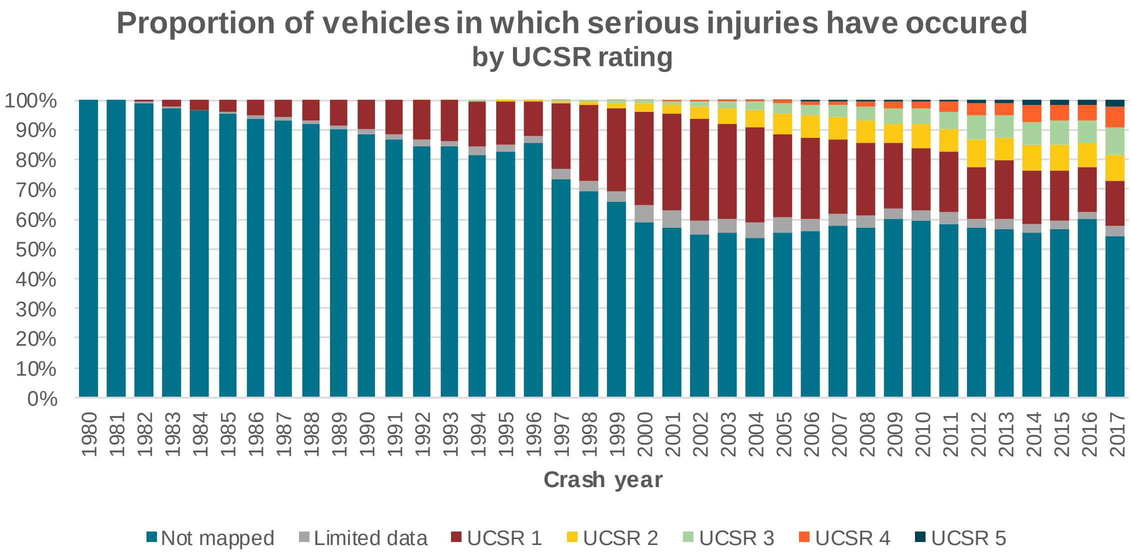
Again these potential factors are considered as part of the multivariate analysis in Section 6.

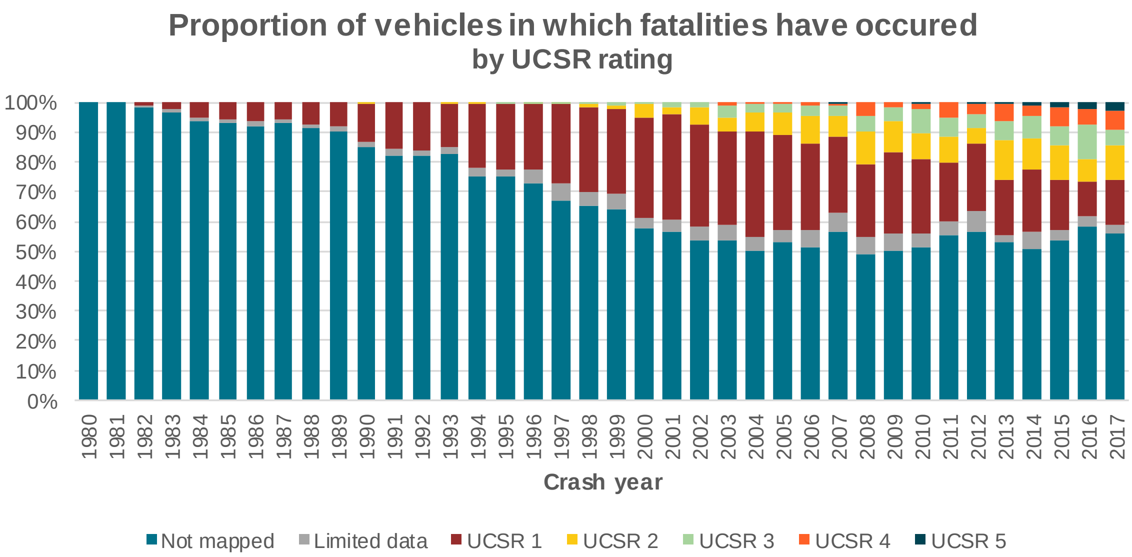




5.5 Fatalities and injuries by UCSR

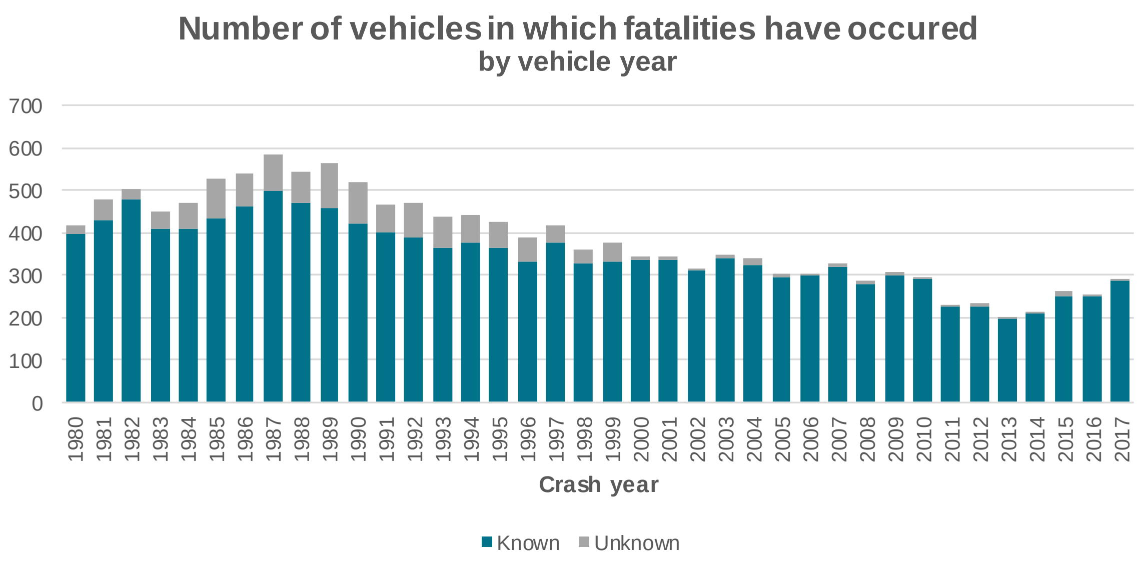
The charts below consider fatalities and injuries by UCSR. In Section 6 we line these figures up against vehicle exposure by UCSR (and simultaneously any other factors that we can) to test the impact of UCSR on fatalities and injuries.





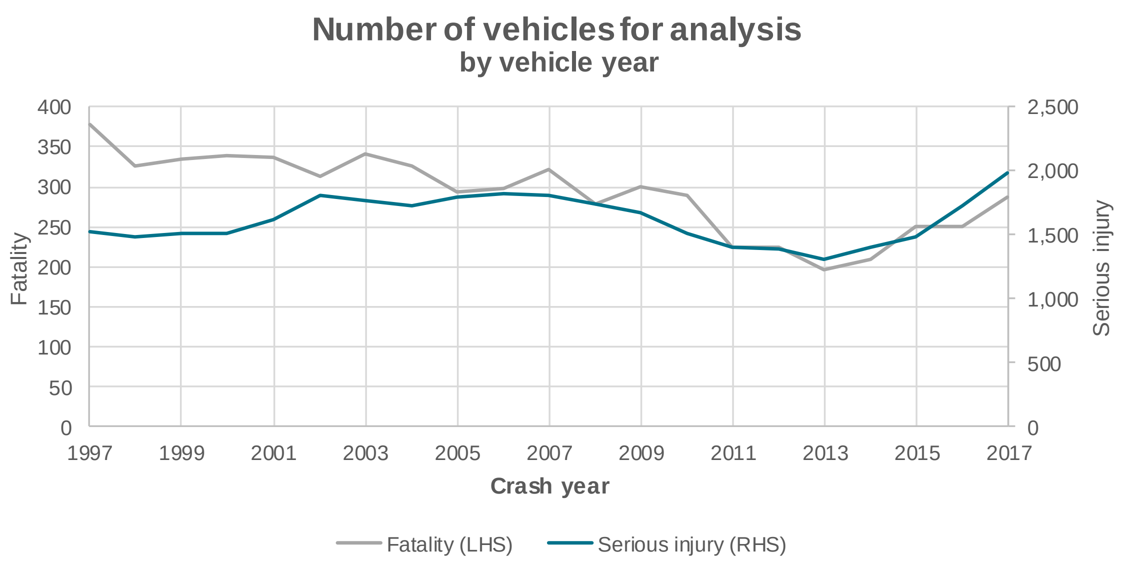
5.6 Accidents in vehicles of unknown age

For a number of records in the CAS dataset the vehicle year had not been recorded. The chart below summarises the fatalities according to whether or not the vehicle year was recorded. The proportions of serious injuries show a similar picture.



If we were able to link accident data to vehicle data using, say, the vehicle license plate, then the problem of unrecorded vehicle year details might reduce. This is discussed in Appendix C. Unfortunately, for this paper, accidents for which the vehicle year has not been recorded are unlikely to be of much use in terms of modelling the factors affecting serious injuries and fatalities. We have therefore removed these accidents from our analysis.

The chart below shows the final serious injury and fatality numbers which were used in the analysis in Section 6. The chart shows only accidents over the period for which we have reliable odometer readings i.e. from 1997 to 2017.



6 Results

In this section we present the results obtained by applying some Generalised Linear Models (GLMs) to the frequency and severity of serious injuries. That is:

- Frequency: the number of vehicles in which serious injuries or fatalities have occurred per billion kilometres travelled
- Severity: the proportion of those accidents which were fatal.

One the of the main purposes of a GLM is to address the combined effect of multiple factors operating in conjunction e.g. to isolate the impact of a vehicle's mass independent of the effect of having a higher or lower ANCAP rating. However, there is one large caveat here: we're only looking at the characteristics of the vehicle. Many of the vehicle characteristics will be correlated with other factors, for example, older cars may generally be driven by younger drivers and for different purposes.

In this paper we have tried to draw as many insights as we can from the vehicle alone, recognising that some of the insights may well be the vehicle acting as a proxy for other factors. In Appendix C we discuss ways in which the analysis could be extended with additional data.

6.1 Frequency

6.1.1 Factors considered

We considered the following factors and whether they might affect the frequency of serious accidents:

- Vehicle year
- Vehicle age at time of accident
- Vehicle mass
- Vehicle class (e.g. motorcycle, passenger car, goods vehicle)
- ANCAP rating
- Year in which ANCAP test was conducted
- Overall ANCAP score (from which the rating was determined)
- Presence of ABS and/or ESC
- UCSR crashworthiness rating
- UCSR crashworthiness score (from which the rating is determined)
- UCSR aggressivity rating
- USCR primary and secondary safety scores

6.1.2 Frequency GLM results

We fitted a GLM with Poisson error structure and log link to the frequency data. The GLM was run iteratively and the impact of removing, reinstating and/or grouping certain variables was tested. The tables on the following page show the final GLM results.

Parameter	Value	Rate per billion km	P-Value
Intercept		9.70	<.0001
UCSR CWR score (%)	0-2	100%	.
	2-3	155%	0.0046
	3-4	192%	<.0001
	4-5	240%	<.0001
	5-6	292%	<.0001
	6-7	356%	<.0001
	7+	381%	<.0001
	Missing	696%	<.0001
UCSR aggressivity rating	4-5	100%	.
	1-3	107%	0.0004
	Missing	102%	0.4762
ANCAP overall rating	5	100%	.
	4	116%	0.0104
	3	122%	0.0119
	1-2	144%	0.0005
	Missing	163%	<.0001
ANCAP rating age	0-4	100%	.
	4-8	108%	0.1852
	8-12	121%	0.001
	12-16	151%	<.0001
	16+	166%	0.0004
	Missing	100%	.
ANCAP ESC	Standard/Optional/Variant	100%	.
	Other	109%	0.109
Vehicle age	0	134%	<.0001
	1-10	100%	.
	10-15	125%	<.0001
	15-20	184%	<.0001
	20+	219%	<.0001

Parameter	Value	Rate per billion km	P-Value
Vehicle year when has no safety rating	2015-2018	100%	.
	2011-2014	117%	0.0381
	2006-2010	203%	<.0001
	2001-2005	165%	<.0001
	1996-2000	140%	<.0001
	1991-1995	130%	0.0001
	1986-1990	123%	0.0027
	1882-1985	106%	0.3916
	Has rating	74%	0.0067
NZTA vehicle class	Passenger vehicle	100%	.
	Moped	1376%	<.0001
	Motorcycle	1066%	<.0001
	Goods vehicle	111%	0.001
	Omnibus	13%	0.0044
	Other	413%	<.0001
Mass when has UCSR CWR score	0-1000	47%	0.1272
	1000-1500	107%	0.0004
	1500-1750	100%	.
	1750-2000	88%	<.0001
	2000-2500	76%	<.0001
	2500-3000	68%	<.0001
	3000-5000	51%	<.0001
	Missing or has no score	22%	<.0001
Mass when has no UCSR CWR score	0-1000	81%	0.0005
	1000-1250	128%	<.0001
	1250-1500	107%	0.0017
	1500-2000	100%	.
	2000-3000	66%	<.0001
	3000-10000	38%	<.0001
	10000-20000	33%	<.0001
	20000+	22%	<.0001
	Missing or has score	95%	0.3156

To interpret the GLM results, the baseline frequency is 9.7 vehicles in serious accidents per billion kilometres. Then, for example, if the vehicle has a UCSR score between 2 and 3 the frequency is multiplied by 155%. If the vehicle also has an aggressivity rating of, say, between 1 and 3 then the frequency again gets multiplied by 107%. And so on for each of the other factors.

Note that a high UCSR 'score' corresponds to a low UCSR crashworthiness star rating i.e. 5 star rated UCSR vehicles are those with the lowest UCSR scores. The UCSR aggressivity rating is a measure of how damaging a vehicle is to other parties (pedestrians, cyclists and other vehicle occupants) in the event of a crash.

The GLM identified that vehicle year is an important factor in determining the serious accident frequency where a vehicle does not have a safety rating. The results by vehicle year may seem somewhat counterintuitive e.g. vehicles manufactured from 2006 to 2010 present a higher risk than vehicles from any other period. In part this is because the vehicle year also needs to be interpreted in the context of the factors for vehicle age (i.e. years between manufacture and accident). However, there also appear to be some interesting effects present by vehicle year/age which are explored in Section 6.1.4.

The presence of ABS was not identified as significant on its own, which is probably because that information is already captured in the ANCAP rating. The UCSR crashworthiness score was identified as more significant than the overall 1-5 rating and also more significant than the primary or secondary safety ratings.

The variables identified by the GLM as being significant effectively fall into three categories:

- The score or rating of the vehicle (UCSR or ANCAP)
- The year and age of the vehicle

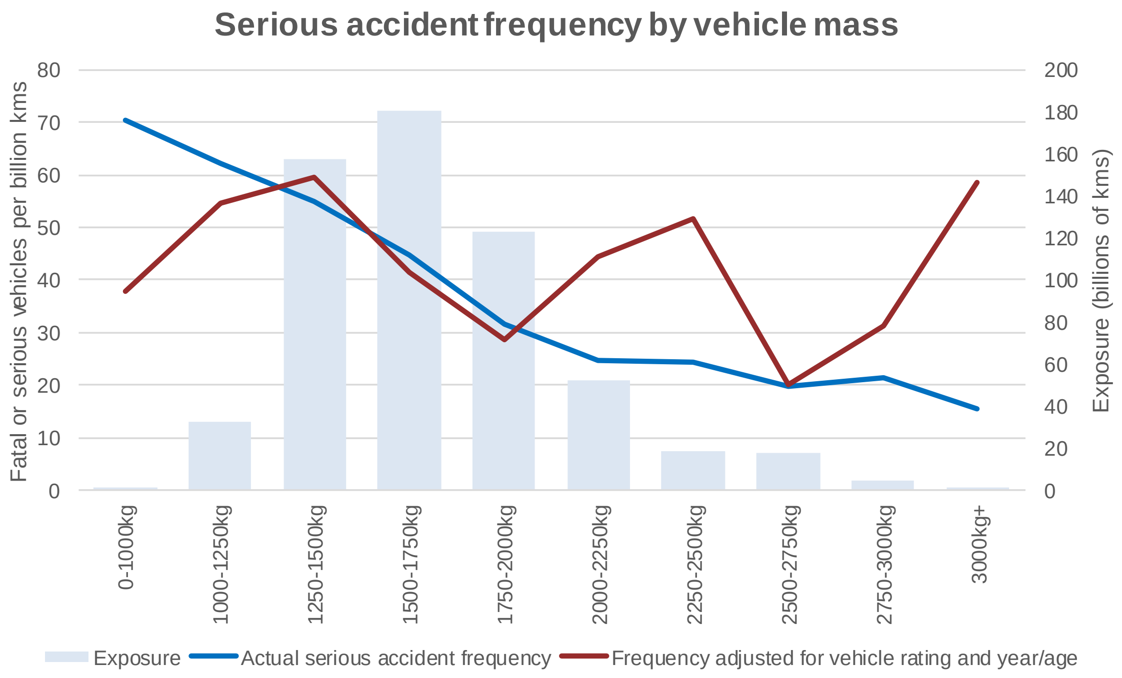
- The class and mass of the vehicle

The vehicle year/age and rating are intimately related i.e. the is a strong correlation between vehicle year and improved safety ratings. Similarly, the mass and vehicle class are closely related.

Rather than discuss every factor in the GLM and what this might mean, we have picked out a few interesting effects to explore. In the sections below we consider separately what the data is telling us in terms of vehicle class/mass and rating/year/age.

6.1.3 Vehicle mass

The chart below demonstrates the impact of vehicle mass on frequency specifically for passenger vehicles.



The blue line shows the effect of vehicle mass based on a simple one way analysis, whilst the red line uses the results of the GLM to adjust for all the other variables. This shows that, whilst there is a strong correlation between vehicle mass and frequency, some of this correlation is removed after allowing for other potential factors. This is because the UCSR scores implicitly allow for vehicle mass to some extent (evidenced by the best GLM result being obtained by having different factors for mass depending on whether or not the vehicle has a UCSR score).

6.1.4 Vehicle year/age and rating

The table below shows the vehicle exposure (i.e. number of kilometres travelled) by vehicle year and age. Each diagonal represents an observation year, for example, the bottom diagonal is travel during 2017. The dark blue areas are combinations of vehicle year and age which contain a lot of exposure (either due to there being lots of vehicles of that combination or vehicles travelling many kilometres). The vehicle age 0 column shows a relatively small amount of exposure because it represents approximately half a year e.g. vehicles first registered in 2013 would, on average, be active for around 6 months during 2013.

Exposure (millions of kms)
All vehicles

Vehicle	Vehicle age (years)																									
year:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1980																		699	531	444	319	235	179	136	104	77
1981																	1,027	788	664	474	355	270	199	142	100	78
1982																1,263	1,003	863	626	467	362	265	194	140	103	76
1983															1,470	1,217	1,082	780	598	455	340	247	175	126	94	68
1984														2,106	1,830	1,696	1,278	1,010	799	610	442	318	223	163	121	95
1985													2,096	1,873	1,803	1,431	1,191	989	791	604	442	320	240	180	139	115
1986												2,096	1,924	1,881	1,525	1,300	1,118	925	726	545	401	310	233	189	154	124
1987											2,299	2,154	2,147	1,788	1,573	1,389	1,192	966	746	560	432	325	263	214	177	136
1988										2,519	2,458	2,520	2,143	1,930	1,749	1,540	1,301	1,039	810	647	499	407	340	276	222	183
1989									2,887	2,889	3,113	2,731	2,539	2,368	2,142	1,866	1,547	1,242	1,026	807	675	568	460	372	303	251
1990								2,616	2,684	3,052	2,903	2,782	2,672	2,466	2,208	1,896	1,581	1,356	1,104	947	818	683	573	475	398	325
1991							1,905	2,035	2,520	2,565	2,665	2,636	2,491	2,277	1,995	1,710	1,503	1,250	1,097	959	806	681	566	468	384	316
1992						1,661	1,884	2,374	2,582	2,759	2,961	2,865	2,697	2,410	2,101	1,874	1,586	1,413	1,248	1,059	896	753	628	512	416	328
1993					1,608	1,695	2,020	2,177	2,416	2,570	2,577	2,466	2,279	2,030	1,842	1,594	1,440	1,292	1,121	973	832	702	588	489	382	
1994				1,823	1,856	2,056	2,215	2,418	2,752	2,889	2,968	2,837	2,626	2,443	2,157	1,994	1,824	1,604	1,420	1,235	1,061	898	754	599		
1995			1,473	1,445	1,556	1,601	1,756	2,056	2,462	2,746	2,928	2,817	2,717	2,444	2,298	2,139	1,913	1,726	1,530	1,337	1,149	982	794			
1996		1,682	1,771	1,785	1,708	1,731	1,990	2,499	3,080	3,422	3,662	3,660	3,450	3,296	3,098	2,828	2,595	2,345	2,101	1,846	1,606	1,329				
1997	652	1,428	1,524	1,479	1,404	1,515	1,818	2,165	2,519	2,740	3,009	2,892	2,795	2,649	2,444	2,262	2,068	1,870	1,664	1,460	1,218					
1998	172	933	1,384	1,303	1,270	1,354	1,535	1,700	1,933	2,152	2,283	2,272	2,169	2,013	1,887	1,745	1,594	1,441	1,290	1,089						
1999	638	1,668	1,570	1,454	1,412	1,468	1,549	1,617	1,795	1,901	2,025	1,964	1,837	1,739	1,628	1,501	1,367	1,242	1,058							
2000	767	1,724	1,658	1,540	1,472	1,466	1,520	1,606	1,709	1,803	1,873	1,856	1,849	1,767	1,654	1,524	1,395	1,187								
2001	741	1,720	1,675	1,569	1,467	1,447	1,489	1,510	1,610	1,761	1,922	1,946	1,863	1,752	1,630	1,507	1,293									
2002	827	1,855	1,799	1,657	1,560	1,549	1,542	1,578	1,723	1,861	1,994	1,934	1,835	1,715	1,606	1,398										
2003	884	2,032	1,949	1,815	1,721	1,656	1,659	1,768	1,895	1,939	1,886	1,795	1,702	1,609	1,417											
2004	953	2,139	2,080	1,955	1,827	1,782	1,866	1,932	2,043	2,238	2,490	2,617	2,741	2,611												
2005	972	2,090	2,040	1,921	1,805	1,783	1,824	1,920	2,166	2,461	2,784	3,041	3,106													
2006	1,038	2,136	2,076	1,955	1,847	1,771	1,787	1,921	2,130	2,389	2,688	2,785														
2007	1,017	2,176	2,135	2,023	1,898	1,840	1,904	2,034	2,223	2,455	2,594															
2008	1,006	2,053	2,012	1,891	1,764	1,724	1,806	1,903	2,019	2,105																
2009	698	1,473	1,406	1,317	1,239	1,225	1,306	1,364	1,375																	
2010	747	1,631	1,593	1,515	1,429	1,425	1,496	1,486																		
2011	820	1,752	1,726	1,637	1,552	1,538	1,553																			
2012	914	1,984	1,961	1,881	1,803	1,723																				
2013	1,005	2,238	2,204	2,141	1,969																					
2014	1,149	2,489	2,335	2,260																						
2015	1,263	2,795	1,899																							
2016	1,484	3,026																								
2017	1,669																									

The are some notable patches of heavy exposure, namely:

- Large number of vehicles at around the 7-14 year age group (likely to be imports)
- Large numbers of vehicles and/or high usage in recent years for new vehicles.

The increasing VKT in recent years can be seen in the chart (the last few diagonals are generally darker than earlier diagonals) but what is interesting is that the increased VKT is not evenly distributed across the diagonal i.e. there appear to be patches of high VKT.

There appears to be an anomaly for vehicles manufactured in 1998 in their ‘zero’ year of travel (i.e. during 1998). This may be due to issues with the early odometer data, hence our analysis is focused more on recent years.

In the next section we look at how this corresponds to numbers of serious injuries and fatalities.

The table below shows the number of serious accident vehicles by vehicle year and age. We have adjusted the figures to allow for variations in vehicle mass and class based on the results of the GLM. The figures have not been adjusted for safety ratings (more on this below).

Number of fatal or serious vehicles - adjusted for mass and vehicle class
All vehicles

Vehicle year:	Vehicle age (years) 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1980																		80	76	46	31	33	30	21	8	7
1981																	127	93	89	62	38	42	22	18	13	10
1982																121	119	104	62	57	43	25	21	17	13	14
1983															125	107	80	79	74	62	46	23	26	21	15	11
1984														131	123	107	117	102	104	78	46	45	42	28	12	11
1985													107	116	110	117	121	112	83	79	73	41	34	27	25	16
1986												129	108	99	89	110	115	84	107	73	52	33	37	33	18	6
1987											91	107	112	128	100	125	122	91	87	64	58	53	31	37	17	15
1988										156	105	121	130	140	158	148	124	117	86	91	51	47	48	48	35	22
1989									131	133	138	144	152	170	183	158	156	122	126	100	96	55	58	41	30	18
1990								112	105	145	149	172	181	167	156	145	142	133	116	117	85	69	62	51	39	49
1991							85	93	101	114	138	149	126	149	135	125	134	120	83	74	69	75	48	47	38	48
1992						51	68	70	114	113	138	157	157	131	141	142	115	123	104	84	83	65	65	46	47	52
1993					33	45	75	77	100	119	131	133	132	114	119	95	71	103	74	83	36	49	64	51	43	
1994				34	49	66	87	110	117	130	124	157	150	129	122	128	104	90	84	92	63	62	94	81		
1995			49	42	43	46	56	68	114	117	145	122	139	127	120	103	102	82	87	78	77	89	104			
1996		57	39	51	59	42	78	91	114	133	198	167	161	154	146	131	121	111	103	109	137	150				
1997	37	48	36	31	46	43	67	76	68	103	104	121	122	113	73	89	74	94	93	124	108					
1998	51	64	45	49	45	59	50	41	83	102	113	98	99	69	74	55	87	88	98	97						
1999	45	61	46	49	57	36	61	76	51	51	82	70	75	47	69	83	63	76	91							
2000	20	62	49	43	34	45	55	67	54	68	66	49	70	59	61	60	81	89								
2001	26	58	44	41	37	31	32	48	66	34	52	49	63	56	62	74	92									
2002	28	54	56	41	53	47	65	56	66	41	67	56	64	65	59	86										
2003	32	72	28	56	53	52	50	34	43	58	44	62	67	65	89											
2004	34	90	73	82	46	57	64	40	53	58	63	78	99	116												
2005	54	91	65	63	49	43	53	62	61	73	91	94	167													
2006	40	77	62	46	54	44	52	46	63	81	81	112														
2007	65	91	45	60	49	53	48	66	64	84	87															
2008	68	100	71	53	55	34	50	38	73	84																
2009	38	37	45	28	30	51	36	36	62																	
2010	32	59	36	37	40	38	40	49																		
2011	25	31	29	23	36	18	39																			
2012	20	36	23	29	40	48																				
2013	24	29	44	39	57																					
2014	32	61	65	54																						
2015	42	48	68																							
2016	47	75																								
2017	46																									

The patches of high injury/fatality numbers broadly correspond to the high VKT, but with the notable exception that the recent high VKT for new vehicles doesn't appear to have resulted in significantly higher accident numbers for these vehicles.

Putting together the exposure and serious accident numbers we get the chart below which shows the serious accident frequency per billion kilometres. We have only shown combinations with a material amount of exposure (in this case at least 200 million kilometres). Note that the proportional impact of increased VKT is already allowed for in this table by calculating accidents per km.

Fatal or serious vehicles per billion kms - adjusted for mass and vehicle class

All vehicles (only showing results with >200 million kms exposure)

Vehicle	Vehicle age (years)																									
year:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1980																		114	143	104	97	142				
1981																	123	118	134	132	106	157				
1982																96	119	121	98	122	119	96				
1983															85	88	74	102	124	135	137	94				
1984														62	67	63	92	101	130	129	103	143	187			
1985													51	62	61	82	102	113	105	131	166	129	141			
1986												62	56	52	58	85	103	91	148	135	129	106	159			
1987											40	50	52	72	63	90	102	95	117	113	133	162	119	175		
1988										62	43	48	61	73	90	96	96	113	106	141	103	116	141	175	157	
1989									45	46	44	53	60	72	86	85	101	98	123	124	143	96	125	110	99	71
1990								43	39	48	51	62	68	68	71	77	90	98	105	124	104	101	108	107	97	152
1991							45	46	40	45	52	56	50	65	68	73	89	96	76	77	86	110	85	101	98	153
1992						31	36	30	44	41	46	55	58	54	67	76	73	87	83	80	93	86	103	90	112	158
1993					20	26	37	35	41	46	51	54	58	56	65	60	50	80	66	85	43	70	108	103	113	
1994				19	26	32	39	45	42	45	42	55	57	53	57	64	57	56	59	74	60	69	124	135		
1995			33	29	27	29	32	33	46	42	50	43	51	52	52	48	53	47	57	58	67	91	132			
1996		34	22	28	35	24	39	36	37	39	54	46	47	47	47	46	47	47	49	59	85	113				
1997	57	34	24	21	32	28	37	35	27	38	35	42	43	43	30	39	36	50	56	85	89					
1998		68	33	38	36	43	32	24	43	47	49	43	46	34	39	31	54	61	76	89						
1999	71	37	29	34	40	24	40	47	28	27	41	36	41	27	43	55	46	61	86							
2000	26	36	29	28	23	31	36	42	31	38	35	26	38	33	37	39	58	75								
2001	35	34	26	26	25	21	21	32	41	19	27	25	34	32	38	49	71									
2002	33	29	31	25	34	31	42	36	38	22	33	29	35	38	37	61										
2003	36	36	15	31	31	32	30	19	23	30	23	34	39	41	63											
2004	35	42	35	42	25	32	34	21	26	26	25	30	36	44												
2005	55	43	32	33	27	24	29	32	28	30	33	31	54													
2006	38	36	30	24	29	25	29	24	30	34	30	40														
2007	64	42	21	30	26	29	25	33	29	34	34															
2008	68	49	35	28	31	20	28	20	36	40																
2009	54	25	32	22	24	41	28	26	45																	
2010	43	36	22	25	28	27	27	33																		
2011	30	18	17	14	23	12	25																			
2012	22	18	12	16	22	28																				
2013	24	13	20	18	29																					
2014	28	25	28	24																						
2015	34	17	36																							
2016	32	25																								
2017	28																									

The table shows two effects:

- Later vehicle years show lower serious accident frequencies.
- Older vehicles (i.e. longer period between manufacture and use) show higher accident frequencies.

The former is largely due to improvements in vehicle safety (identified as a significant factor – see the GLM results table) whilst the latter is possibly due to older vehicles being driven by younger drivers. In Section 6.1.2 we quantified the impact on frequency of having a higher UCSR score or ANCAP rating. On the following page we take a look at vehicles which don't have a UCSR or ANCAP rating.

The table below looks at serious accident frequencies for vehicles without a UCSR or ANCAP rating.

Fatal or serious vehicles per billion kms - adjusted for mass and vehicle class
Vehicles without a UCSR or ANCAP rating (only showing results with >100 million kms exposure)

Vehicle year:	Vehicle age (years) 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1980																		104	131	95	88	130	154	142	67	
1981																	113	109	122	121	98	141	103	115		
1982																104	114	121	110	113	143	143	154			
1983															83	92	90	94	119	146	152	98				
1984														70	69	66	83	90	119	112	118	154				
1985													60	71	69	81	103	102	120	103	150	122	144			
1986												69	76	63	67	85	92	80	122	133	130	95	149			
1987											43	59	58	73	68	79	75	97	114	114	86	183	126	172	86	
1988										71	54	52	66	76	83	111	86	115	101	143	106	98	160	172	195	
1989									51	54	47	53	64	71	89	81	109	104	149	118	167	114	139	143	99	50
1990								56	42	57	62	58	76	77	68	70	107	105	95	125	103	106	113	115	96	150
1991							61	57	46	52	64	64	40	59	64	69	89	88	72	72	104	114	72	125	127	172
1992						41	53	30	53	45	55	61	67	60	74	74	77	92	93	83	94	89	116	116	112	141
1993					28	32	46	33	52	43	57	50	60	58	65	67	57	83	64	76	46	67	98	90	92	
1994				27	38	34	43	51	42	43	44	54	62	54	49	72	69	63	61	61	52	99	142	148		
1995			43	42	38	31	32	36	48	47	55	44	53	52	65	48	65	52	52	59	62	101	100			
1996		61	29	53	45	41	49	42	41	43	54	56	44	55	44	39	56	54	47	56	88	117				
1997	111	60	38	46	57	42	62	48	35	49	39	48	41	53	41	52	41	57	46	105	93					
1998		112	71	77	69	69	53	41	57	58	54	49	56	41	46	39	49	74	109	102						
1999	208	88	71	84	93	58	71	70	59	36	51	53	60	25	43	51	58	88	126							
2000	60	62	52	63	54	57	61	63	63	72	57	45	55	52	48	52	88	74								
2001	74	66	53	50	62	65	49	80	94	29	44	44	58	54	47	72	77									
2002	68	67	65	52	76	72	96	85	82	45	49	60	63	54	48	83										
2003	100	89	49	71	95	85	76	57	53	63	56	63	86	78	102											
2004	124	116	95	97	66	90	80	60	61	55	42	54	53	58												
2005	182	143	119	110	101	72	76	99	61	68	56	51	74													
2006	150	135	113	102	81	76	89	64	88	77	62	67														
2007	245	162	96	127	102	115	107	95	90	92	76															
2008	280	180	152	115	117	77	82	67	112	105																
2009	225	117	125	92	109	155	94	100	116																	
2010	187	158	90	102	108	90	96	91																		
2011	135	61	63	63	62	53	57																			
2012	88	56	47	61	76	92																				
2013	90	50	70	78	113																					
2014	86	72	95	61																						
2015	112	59	92																							
2016	111	69																								
2017	73																									

The table highlights some interesting features:

- High accident frequencies for vehicles manufactured around 2005 to 2010 in their first few years of use.
- Higher frequencies for early model and/or older vehicles (although the distinction is perhaps not as strong for unrated vehicles as for vehicles as a whole).
- The most recent one or two diagonals are generally more towards the high frequency end than earlier diagonals (i.e. the recent increase in serious injury numbers discussed in Section 5).

The GLM identified unrated vehicles manufactured from 2006 to 2010 as high risk – the table above illustrates why the GLM has identified these vehicles in particular. Most vehicle years show increased frequencies in the last diagonal. However, the frequencies for unrated vehicles manufactured around 2008 to 2009 are particularly high – comparable to that for unrated vehicles prior to 1999.

As one further drilldown we consider the exposure and frequency specifically for motorcycles (given that increased motorcycle registrations were identified by Deloitte as a contributor to the increased road toll in recent years). The tables on the following page show the exposure (VKT) and serious accident frequency for motorcycles (all of which are unrated).

Exposure (millions of kms)

Motorcycles

Vehicle year:	Vehicle age (years) 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1990								5.7	4.6	5.6	5.0	4.1	4.0	3.7	3.6	3.1	3.4	3.2	2.6	2.7	2.4	2.3	2.2	2.2	1.7	1.5
1991							4.6	3.6	3.7	3.2	3.7	3.7	3.2	2.6	3.4	2.9	2.8	2.6	2.3	2.0	1.8	1.5	1.7	1.5	1.5	1.3
1992						2.5	2.2	2.4	2.5	2.1	2.6	2.0	1.8	2.0	2.0	2.3	2.1	2.4	1.8	1.7	1.5	1.4	1.5	1.1	0.9	0.7
1993					2.4	1.7	2.0	2.2	1.8	1.6	1.7	1.6	1.7	1.8	1.7	2.1	1.8	1.9	1.5	1.2	1.2	1.2	0.9	1.0	0.8	
1994				2.3	2.3	2.3	2.3	2.1	2.3	1.8	1.8	2.3	2.5	2.4	2.0	1.9	2.1	1.8	1.5	1.5	1.4	1.8	1.3	1.4		
1995			3.5	3.7	4.0	3.4	3.9	3.8	3.1	2.7	3.0	3.2	3.6	2.9	2.9	3.4	2.9	3.1	2.2	1.8	2.1	2.2	2.1			
1996		5.9	5.8	5.6	5.1	4.4	4.1	3.6	4.5	3.5	3.7	3.2	3.2	3.2	3.0	2.8	2.7	2.6	2.2	2.3	2.1	1.4				
1997	3.6	5.7	5.3	5.1	5.0	4.1	3.9	4.1	4.3	3.9	3.9	3.5	3.3	2.7	2.4	2.6	2.7	2.2	2.0	2.2	1.9					
1998	0.5	2.9	4.8	4.7	4.3	4.0	3.9	4.1	3.7	3.9	3.7	4.2	3.1	2.7	2.6	2.7	2.4	2.1	1.8	1.9						
1999	2.9	6.0	5.2	4.7	4.2	4.1	4.2	4.3	4.3	4.2	3.9	4.0	3.3	3.6	3.7	2.9	2.6	2.4	2.2							
2000	3.4	5.5	4.6	4.2	3.9	3.9	4.1	4.1	4.3	4.1	3.6	3.3	3.2	3.2	2.7	2.8	2.5	2.1								
2001	3.5	6.5	5.8	5.5	5.5	5.8	5.8	5.8	5.2	5.0	4.5	4.3	4.4	3.8	3.6	3.3	2.9									
2002	4.0	7.3	6.8	6.5	6.4	6.3	6.2	6.0	5.4	4.8	4.7	4.5	4.1	4.1	3.8	3.3										
2003	4.3	7.7	7.5	7.6	7.3	7.2	6.7	6.0	5.4	5.1	4.9	4.5	4.6	4.7	4.2											
2004	5.4	10.9	10.4	9.9	9.5	8.6	7.9	7.0	6.4	6.3	6.3	5.6	5.3	4.9												
2005	6.8	13.3	12.6	11.6	10.4	9.6	9.1	9.0	8.5	8.3	7.7	7.5	6.4													
2006	7.1	11.8	11.3	10.2	9.1	8.5	8.0	7.7	7.2	6.8	6.9	6.1														
2007	13.3	23.2	20.6	18.1	16.3	15.3	14.3	13.4	12.6	11.6	10.8															
2008	11.3	20.8	18.2	16.2	15.0	14.1	13.1	12.1	12.0	10.2																
2009	8.5	14.4	12.3	11.0	10.4	9.6	9.1	8.8	8.2																	
2010	6.3	10.6	9.7	8.7	8.2	7.6	7.0	6.1																		
2011	4.8	9.3	8.6	8.0	7.4	7.1	6.4																			
2012	4.9	9.3	8.4	7.9	7.3	6.6																				
2013	5.1	10.6	9.6	8.9	7.7																					
2014	6.7	13.9	9.4	9.7																						
2015	16.3	36.8	13.1																							
2016	24.9	45.5																								
2017	26.2																									

Fatal or serious vehicles per billion kms - adjusted for vehicle mass

Motorcycles (only showing results with >1 million kms exposure)

Vehicle year:	Vehicle age (years) 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1990								532	663	537	603	748	1,017	1,928	859	1,004	290	641	379	360	867	428	-21	464	580	674
1991							436	841	264	1,607	262	261	302	1,580	290	-21	357	773	428	498	573	-22	-22	-21	1,382	811
1992						404	456	853	1,230	466	785	498	2,247	504	501	-22	468	411	551	-21	-22	733	688	1,921		
1993					1,300	602	511	949	-22	-22	-21	-22	1,802	564	-22	478	-22	519	652	836	-21	-21		-21		
1994				-20	871	-21	892	476	-21	553	567	-21	398	-21	-21	518	962	-22	664	-21	-21	547	758	726		
1995			569	-20	-20	600	-20	251	-20	751	682	306	-21	-21	710	-21	340	-21	-21	560	467	1,899	991			
1996		509	-21	537	800	214	237	1,143	-20	276	-21	-21	-21	311	1,040	352	-21	381	-21	892	1,490	3,591				
1997	838	527	-21	389	809	232	520	745	462	1,587	515	274	294	1,509	-21	1,951	361	913	1,028	2,330	3,830					
1998		1,074	194	426	467	-21	244	-21	-21	510	540	226	312	363	-21	367	-21	486	2,240	1,061						
1999	704	325	381	644	470	1,005	477	468	949	225	-21	240	294	-21	-21	-21	384	411	2,363							
2000	587	167	208	720	1,330	244	-21	738	224	-21	1,722	-21	304	309	-21	725	1,624	974								
2001	276	140	159	742	357	-21	159	340	585	401	681	462	218	258	553	1,562	2,521									
2002	241	548	443	140	141	-21	316	329	751	-21	-21	-21	232	236	813	609										
2003	465	1,198	258	392	263	268	449	675	365	385	193	681	435	425	1,477											
2004	750	841	-21	294	89	828	111	279	465	145	144	-21	769	1,699												
2005	592	370	476	428	583	524	324	441	345	607	385	817	1,940													
2006	274	67	71	697	94	102	758	251	270	592	1,650	838														
2007	449	474	29	267	235	115	418	135	311	1,239	1,527															
2008	625	480	151	364	397	127	298	236	852	1,921																
2009	225	341	571	73	379	523	322	455	1,131																	
2010	645	-21	411	218	359	526	874	1,003																		
2011	196	202	100	110	260	566	467																			
2012	404	-22	227	110	550	1,089																				
2013	182	77	196	449	1,070																					
2014	135	129	866	729																						
2015	234	92	1,091																							
2016	399	277																								
2017	178																									

The increase in exposure to late model motorcycles in recent years is clearly evident. However on a frequency basis there is less evidence of an upsurge. In other words, the increase in serious injuries and fatalities for late model motorcycles is perhaps roughly proportional to the increased exposure for say 2016 and 2017 manufactured motorcycles.

However, for older motorcycles there is clearly an increase in serious injuries and fatalities beyond that expected from the increase in exposure i.e. the frequency per kilometre has increased for the latest diagonals on the table above. This may be related to driver characteristics e.g. younger drivers riding older motorcycles.

This high accident frequency for unrated vehicles of a particular era, including older motorcycles, is worth investigating further. However, at this point an understanding of other factors (e.g. the driver or accident contributors) is necessary in order to truly understand what is causing this increase.

In the next section we consider the factors affecting accident severity.

6.2 Severity

6.2.1 Factors considered

To model severity we considered all the same factors as we did for frequency with the addition of:

- The mass of the largest other vehicle in the accident
- The ratio of the masses of the vehicles in an accident
- The aggressivity rating of the most aggressive other vehicle
- The number of other vehicles involved (if any) and the type of other vehicle (e.g. motorcycle, passenger car, truck)

6.2.2 Severity GLM results

We fitted a GLM with Poisson error structure and log link to the severity data. Similarly to the frequency model, the GLM was run iteratively and the impact of removing, reinstating and/or grouping certain variables was tested. The tables below show the final GLM results.

Parameter	Value	Rate	P-Value
Intercept		6.88%	<.0001
UCSR crashworthiness rating	5	100%	.
	4	120%	0.4188
	3	128%	0.2539
	2	143%	0.0994
	1	147%	0.0723
	Missing	152%	0.0568
UCSR aggressivity rating	3-5	100%	.
	2	111%	0.038
	1	134%	<.0001
	Missing	110%	0.4421
Vehicle year when has no UCSR rating	2011-2017	100%	.
	2006-2010	112%	0.3857
	2001-2005	132%	0.0226
	1996-2000	142%	0.003
	1986-1995	158%	<.0001
	1981-1985	166%	<.0001
	1902-1980	181%	<.0001
	Has rating	101%	0.9574
NZTA vehicle class	Passenger car/forward control passenger vehicle/bus	100%	.
	Off-road passenger vehicle	80%	0.0025
	Goods vehicle	74%	<.0001
	Motorcycle	80%	0.0488
	Moped	35%	0.0258
	Other	69%	0.0018

Parameter	Value	Rate	P-Value
Mass when has no UCSR rating	0-1000	110%	0.5512
	1000-2000	100%	.
	2000-3000	119%	0.0246
	3000-10000	105%	0.7182
	10000+	208%	<.0001
	Missing or has score	156%	0.0133
Other vehicles involved	None involved	100%	.
	Others involved	86%	0.02
Other cars involved	None involved	100%	.
	Others involved	86%	0.0094
Other vans involved	None involved	100%	.
	Others involved	128%	0.0002
Other motorcycles	None involved	100%	.
	Others involved	141%	0.0054
Other SUVs involved	None involved	100%	.
	Others involved	127%	0.0018
Other trucks involved	None involved	100%	.
	Others involved	242%	<.0001
Other buses involved	None involved	100%	.
	Others involved	154%	0.0007
Mass of heaviest other vehicle	0-1000	65%	0.46
	1000-1500	84%	0.0343
	1500-2000	100%	.
	2000-3000	106%	0.4388
	3000-10000	75%	0.0179
	10000+	83%	0.1906
	Missing or has score	98%	0.7595
Ratio of mass of heaviest other vehicle to this vehicle	0-0.75	72%	0.0206
	0.75-5	100%	.
	5+	127%	0.079
	Missing or has score	99%	0.8909

To interpret the GLM results, the baseline severity is a 6.88% probability that a vehicle in which there is at least a serious injury will indeed include a fatality. Then, for example, if that vehicle has a UCSR crashworthiness rating of 1 the severity is multiplied by 147%. A contrast to the frequency model is that for the severity model we are also able to consider the nature of the other vehicle(s) involved in the crash.

The factors identified broadly fall into four categories:

- The UCSR crashworthiness and aggressivity rating for the primary vehicle
- The vehicle year for the primary vehicle (if there is no UCSR rating)

- The mass and type of the primary vehicle
- The mass and type of the other vehicle(s)

The aggressivity of the other vehicle(s) was not identified as significant by the GLM. This could be because:

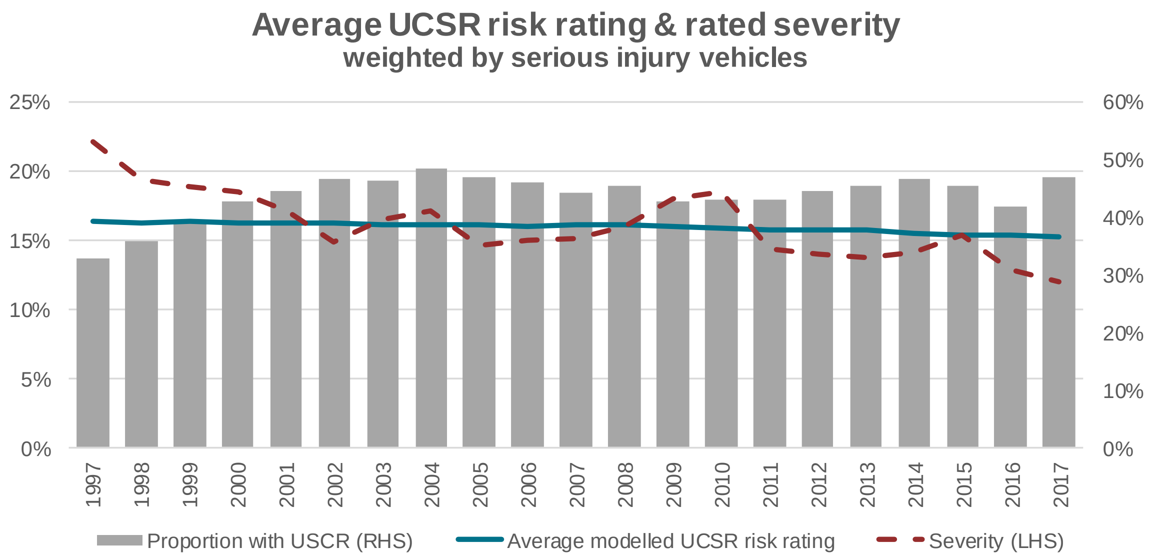
- The aggressivity rating also considers injury to pedestrians and/or cyclists, which are not part of this analysis.
- The aggressivity rating is correlated with the vehicle mass, and the vehicle mass proved to have greater explanatory power given that the vehicle mass data was more complete.

In the sections below we look into the impacts of UCSR and vehicle year, as well as considering the other vehicle involved.

6.2.3 UCSR

Where UCSR ratings were available the GLM identified both the crashworthiness rating and the aggressivity rating as predictive of whether or not a serious accident would be fatal. That is, vehicles with a better crashworthiness rating or with a worse aggressive rating are less likely to see a serious injuries become a fatality. It would appear that, to some extent, more aggressive vehicles protect their occupants at the expense of other road users.

In the chart below we have plotted the overall UCSR effect over time i.e. the weighted average GLM factor for crashworthiness and aggressivity by crash year. This is compared to the observed portion of serious accidents resulting in fatality. Also shown is the proportion of serious accidents in vehicles with a USCR rating at all (which has been broadly similar since the early 2000s).

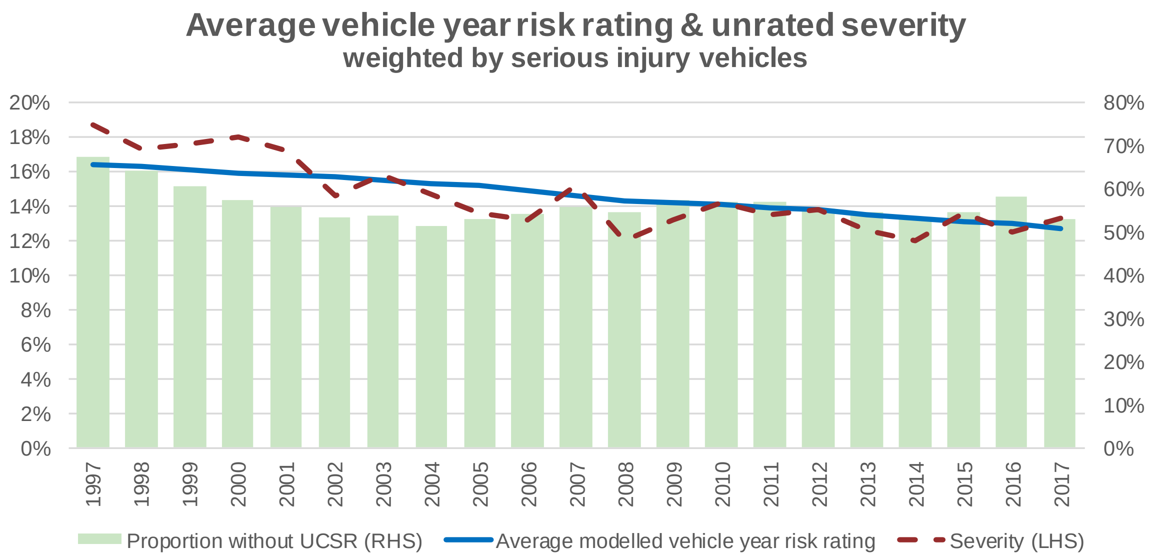


The generally decreasing severity over time is correlated with the decrease in riskiness according to the UCSR crashworthiness and aggressivity ratings. However, the decrease in vehicle riskiness is very gradual. This means that, although the GLM identified the UCSR ratings as an important predictor of severity, the proportion of serious accidents involving better UCSR rated vehicles over time is only changing very gradually. This might seem at odds with the charts in Section 4.4 which show materially improving UCSR ratings over time. However, the frequency analysis also identified UCSR scores as an important factor, so an increasing number of better rated vehicles won't feature in the severity analysis as they avoided serious injury altogether.

6.2.4 *Vehicle year*

Where UCSR ratings were not available the GLM identified the vehicle year as predictive of whether or not a serious accident would be fatal. Later model vehicles, by and large, are less likely to see a serious injury become a fatality.

In the chart below we have plotted the overall vehicle year effect over time i.e. the weighted average GLM factor for vehicle year by crash year. This is compared to the observed portion of serious accidents resulting in fatality. Also shown is the proportion of serious accidents in vehicles with no USCR rating.



The decreasing severity for non-UCSR rated vehicles over time is correlated with improvements in riskiness according to vehicle year. The improvements over time for unrated vehicles show a stronger trend than that for UCSR rated vehicles as shown in Section 6.2.3 above.

The charts in this and the previous section serve to illustrate what the GLM has identified i.e. that reductions in severity over time are a function of:

- Improved UCSR ratings of vehicles involved in accidents
- Later model non-UCSR rated vehicles involved in accidents.

In the next section we consider the impact of the other vehicle(s) involved in an accident.

6.2.5 Other vehicles

Where the primary vehicle (i.e. the vehicle for which we are assessing severity) does not have a UCSR rating, the GLM identified the mass of the other vehicle(s) as predictive of severity. The table below shows the number of primary vehicle serious or fatal accidents according to:

- The mass of the largest other vehicle
- The ratio of the mass of the primary vehicle to the largest other vehicle
- The type of other vehicle(s) involved.

Multiple vehicle accidents in which the primary vehicle does not have a UCSR
Number of vehicles in which serious injury or fatality have occurred

Mass of the heaviest other vehicle	Vehicle mass ratio	Accidents involving other: Motorcycles	Cars	Vans	SUVs	Trucks	Buses
0-1,000kg	0-75%	24	1	2	-	1	-
1,000-1,500kg	0-75%	16	1,597	8	6	16	3
1,500-2,000kg	0-75%	49	1,933	62	74	22	6
2,000-3,000kg	0-75%	32	382	578	314	17	3
3,000-10,000kg	0-75%	3	27	79	15	111	9
10,000kg+	0-75%	6	12	4	1	131	5
0-1000	75%-500%	5	-	-	-	-	-
1000-1500	75%-500%	1	801	5	6	14	3
1500-2000	75%-500%	5	1,658	53	47	19	3
2000-3000	75%-500%	6	401	650	332	27	5
3000-10000	75%-500%	-	45	92	10	130	8
10000+	75%-500%	-	3	-	-	22	-
0-1000	500%+	-	-	-	-	-	-
1000-1500	500%+	1	66	1	-	-	-
1500-2000	500%+	11	121	2	4	-	1
2000-3000	500%+	3	34	37	12	1	1
3000-10000	500%+	1	13	10	8	62	7
10000+	500%+	-	23	5	3	192	7

The dark blue combinations are those with the largest number of serious accidents. Some combinations are very rare or impossible e.g. accidents involving trucks of very small mass. The purpose of the table is to highlight the important combinations from a severity perspective.

On the following page we show the various combinations of factors identified by the GLM as predictive of severity.

Multiple vehicle accidents in which the primary vehicle does not have a UCSR
Relative risk identified by the GLM for severity

Mass of the heaviest other vehicle	Vehicle mass ratio	Accidents involving other: Motorcycles	Cars	Vans	SUVs	Trucks	Buses
0-1,000kg	0-75%	164%	-	149%	-	-	-
1,000-1,500kg	0-75%	214%	130%	194%	192%	367%	234%
1,500-2,000kg	0-75%	254%	155%	231%	229%	437%	278%
2,000-3,000kg	0-75%	270%	164%	245%	243%	464%	296%
3,000-10,000kg	0-75%	192%	117%	174%	173%	330%	210%
10,000kg+	0-75%	211%	128%	191%	-	362%	230%
0-1000	75%-500%	229%	-	-	-	-	-
1000-1500	75%-500%	-	182%	271%	268%	512%	326%
1500-2000	75%-500%	355%	216%	322%	319%	609%	388%
2000-3000	75%-500%	377%	229%	342%	339%	647%	412%
3000-10000	75%-500%	-	163%	243%	241%	459%	293%
10000+	75%-500%	-	179%	-	-	504%	-
0-1000	500%+	-	-	-	-	-	-
1000-1500	500%+	-	231%	-	-	-	-
1500-2000	500%+	450%	274%	409%	405%	-	-
2000-3000	500%+	478%	291%	434%	430%	-	-
3000-10000	500%+	-	207%	308%	305%	583%	372%
10000+	500%+	-	227%	338%	335%	640%	408%

As might be expected, the accidents which are most likely to result in a fatality are those involving large vehicles and/or with a high secondary/primary vehicle mass ratio.

One, perhaps counterintuitive, finding is that the primary vehicle severity for serious accidents where the secondary vehicle is a motorcycle is lower than that where the secondary vehicle is a car. This is probably because we are only looking at accidents where the primary vehicle incurred at least a serious injury. Many accidents involving motorcycles will result in no or minimal injury to the primary vehicle driver or passengers. By narrowing the severity analysis to only those accidents in which the primary vehicle occupants were seriously injured we are perhaps looking at a unique category of accident e.g. where the primary vehicle left the road at high speed after colliding with a motorcycle.

Another apparently odd result is that severity decreases slightly as the other vehicle mass exceeds 3,000kgs. This needs to be understood in the context of the other variables. For example, where other very large vehicles are involved, the severity impact is perhaps reflected more in the vehicle mass ratio than the overall vehicle mass, or by the fact that the accident will necessarily involve a truck, and it is the truck severity factor which drives the result.

As with the frequency analysis, the results of the severity analysis would be worth investigating further with a better understanding of non-vehicle factors.

NZ Society of Actuaries Conference
October 2018

An actuarial approach to modelling vehicle injuries and fatalities

Appendices

MELVILLE JESSUP WEAVER

Willis Towers Watson Alliance Partner

A Data supplied

A.1 Motor vehicle registry

Extract supplied by the NZTA, containing a history of odometer readings for 7.7 million registered vehicles. The fields included a unique (anonymous) Vehicle ID, make, and model of the vehicle. The data contained around 114 million records. The readings were recorded from 1995 to 2018.

We also utilised the NZ open fleet data which gives details of every currently registered vehicle in the country to obtain information on vehicle mass and categorisation.

A.2 Casualty Analysis System

Extract from the CAS showing all vehicles accidents in which injury or fatality occurred from 1980 to 2018. The fields included (anonymised) Crash ID, vehicle counter, vehicle make and model, number of people with fatality, serious injury, minor injury.

A.3 ANCAP ratings

Spreadsheet showing ANCAP data for a number of vehicle year/make/model combinations. Data included ANCAP scores, star ratings, information on various features like ABS and ESC.

A.4 Used Car Safety Ratings

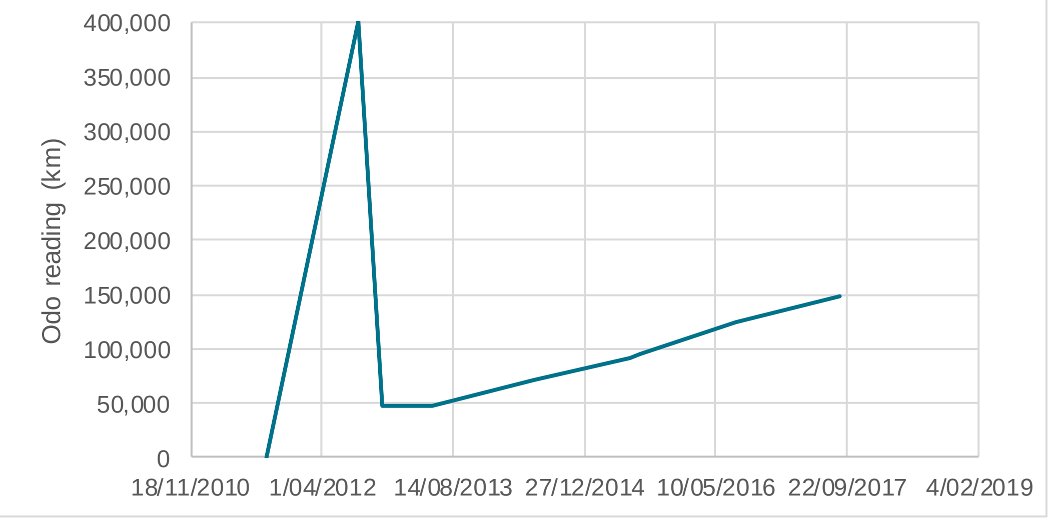
Spreadsheet showing used car safety information for a number of vehicle year/make/model combinations. Data included crashworthiness score and star rating, primary/secondary safety score and aggressivity rating.

B Mileage readings – data cleaning

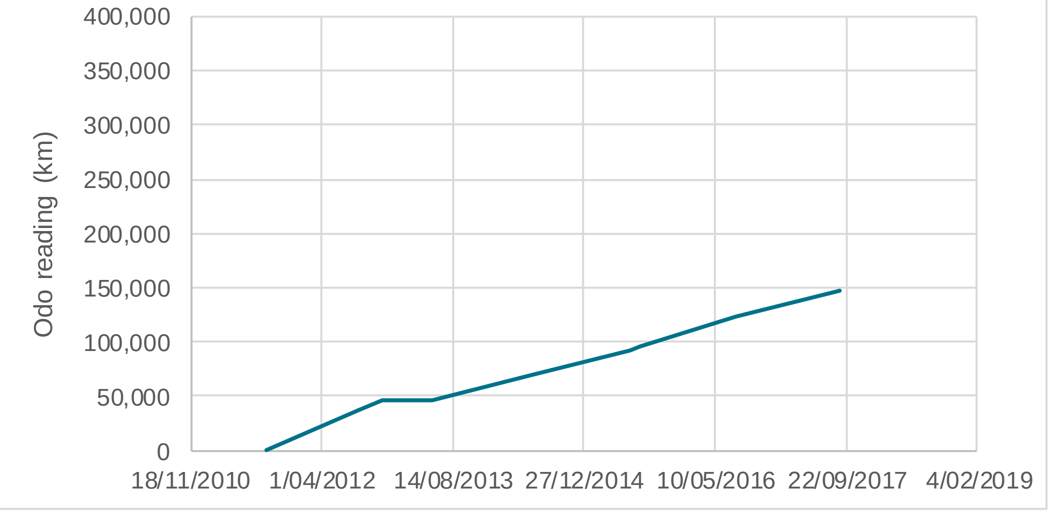
A number of irregularities were apparent in the MVR odometer extract which required remedying by way of smoothing and projecting. Irregularities were defined as readings which moved backwards. Common cases and their fixes are detailed below.

B.1 Single spikes

Clear recording errors due to an extra digit being inputted.

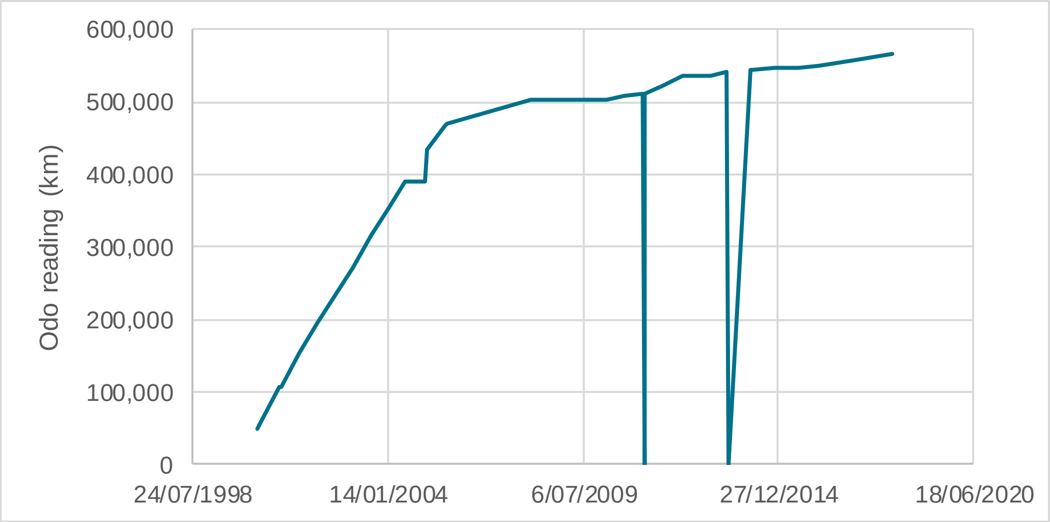


These were smoothed from the values on either side.

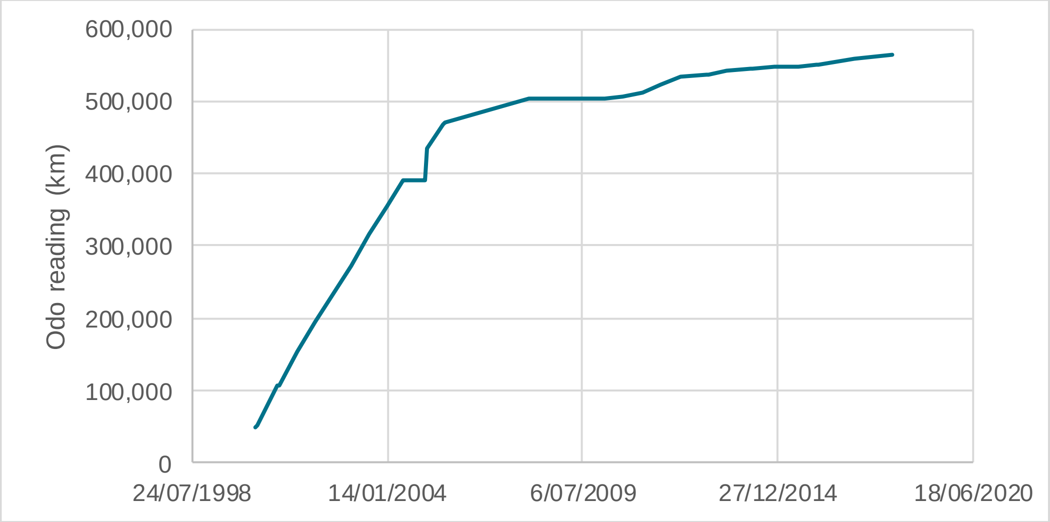


B.2 Missing records

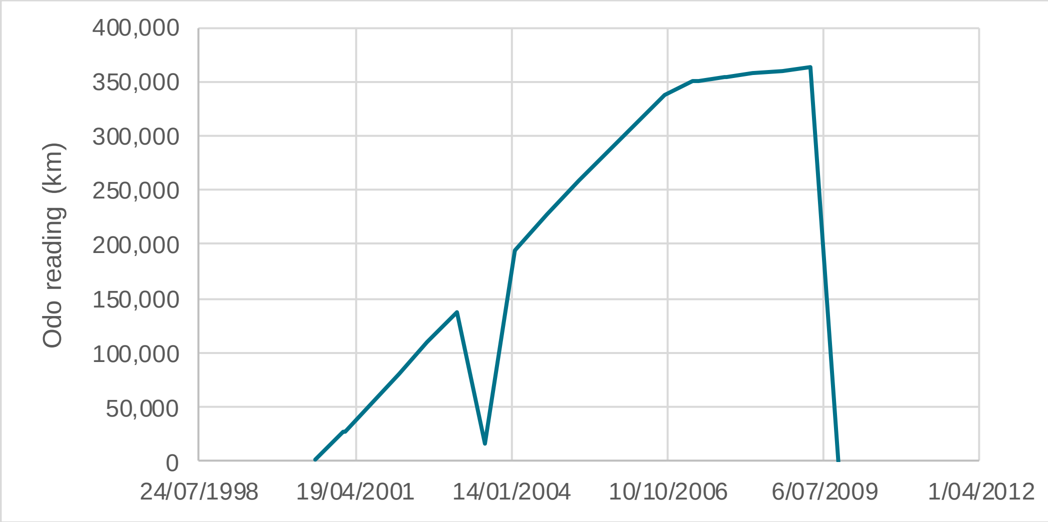
These were zero or near-zero readings which occur in the middle of a vehicle history or at the end.

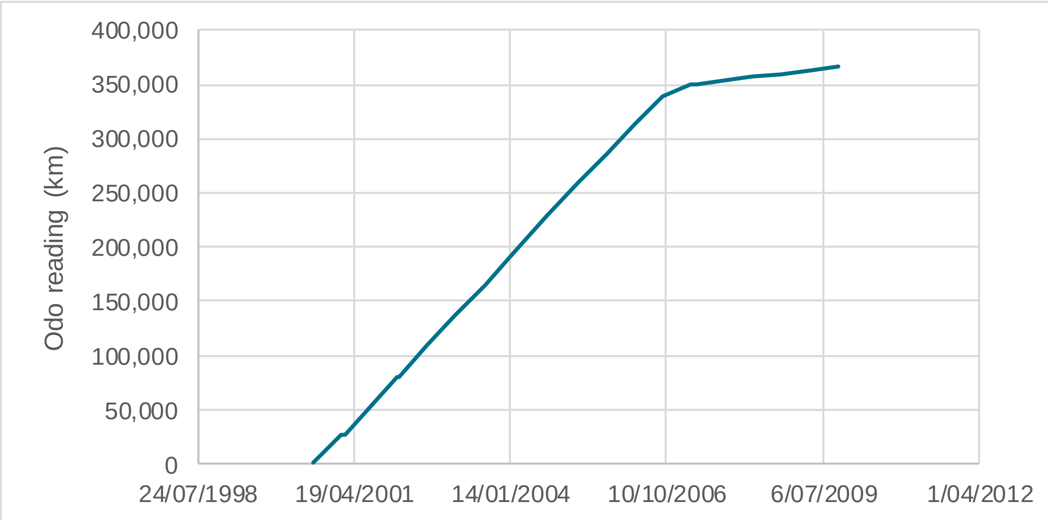


Missing entries with valid readings available on either side were smoothed over.



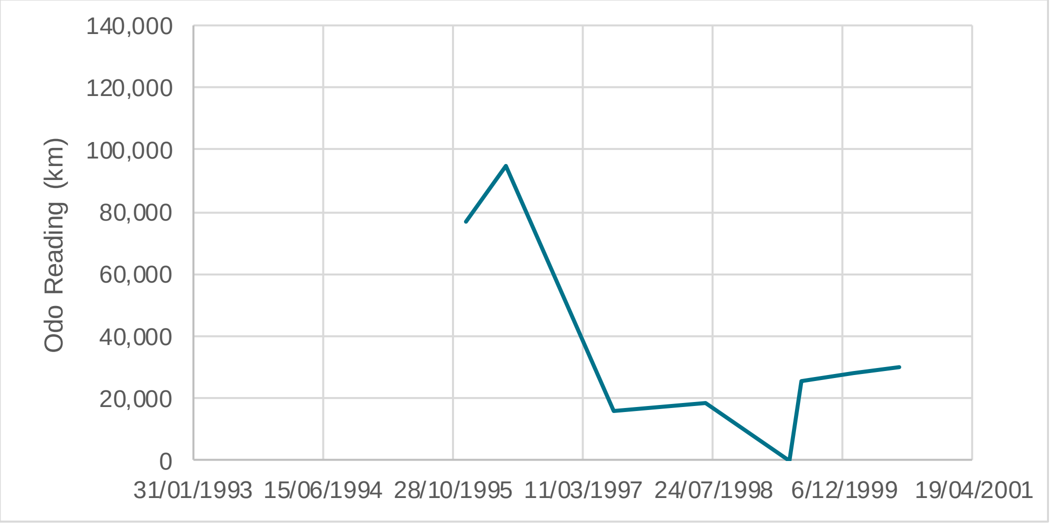
Terminal missing entries were projected forward from the previous reading.



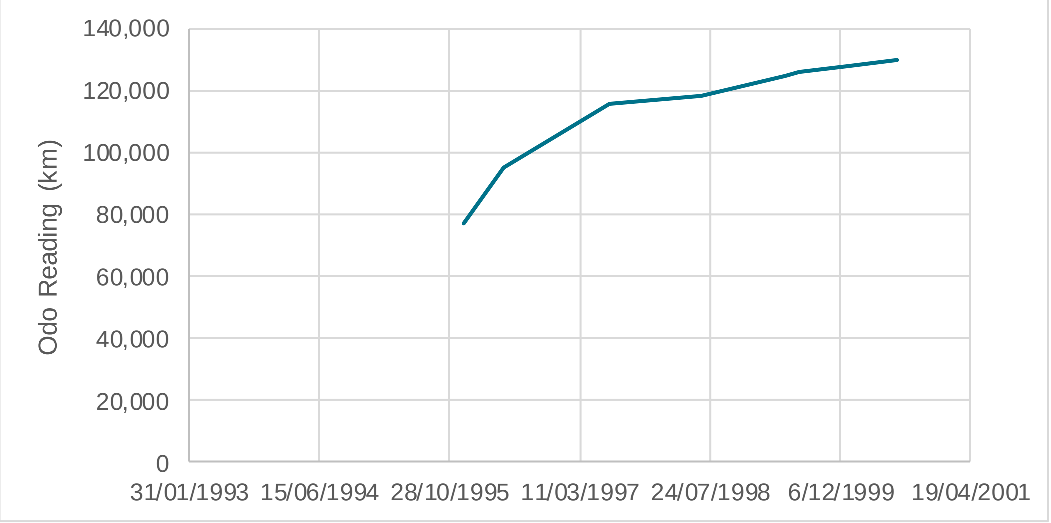


B.3 Odometer tick-overs

Older model vehicles with 5-digit odometers which ticked over 99,999 kms would show a large drop off at the next reading.



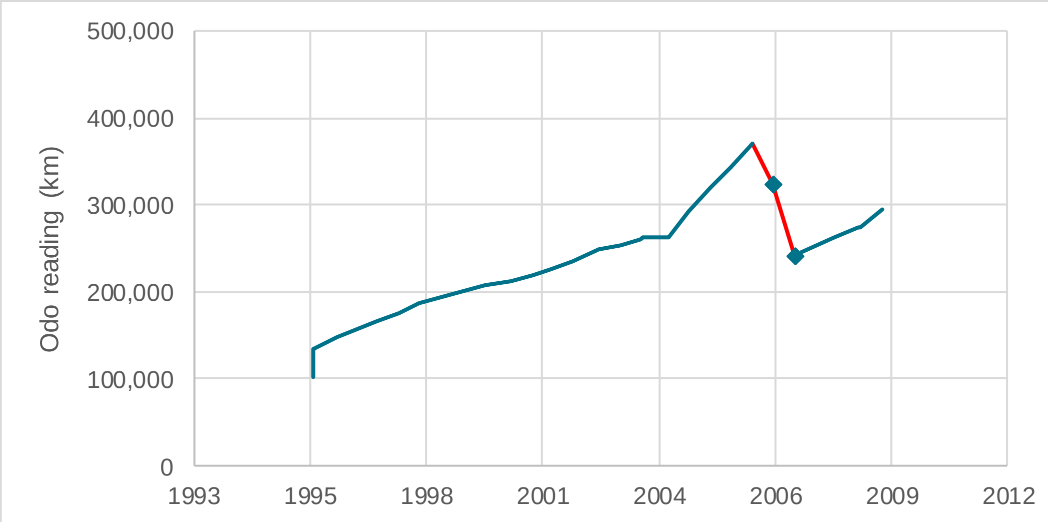
All readings following the presumed tick-over event are increased by 99,999. Note this example also shows a missing reading between 1998-1999 which was smoothed over.



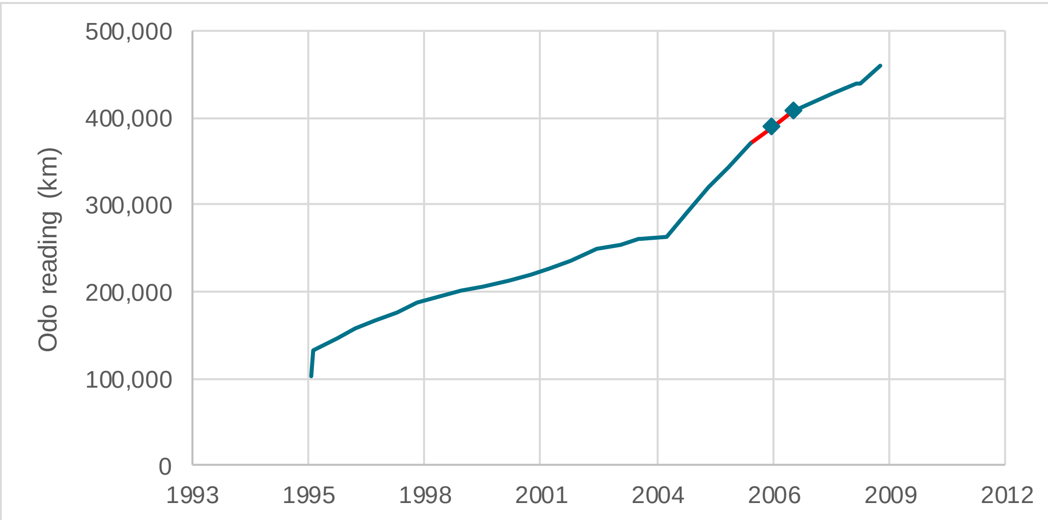
There was also a small population of likely 6-digit odometer tick-overs and these were treated in a similar fashion.

B.4 Consecutive runs of negatives

Some vehicles showed a string of decreases before the readings start to rise again.

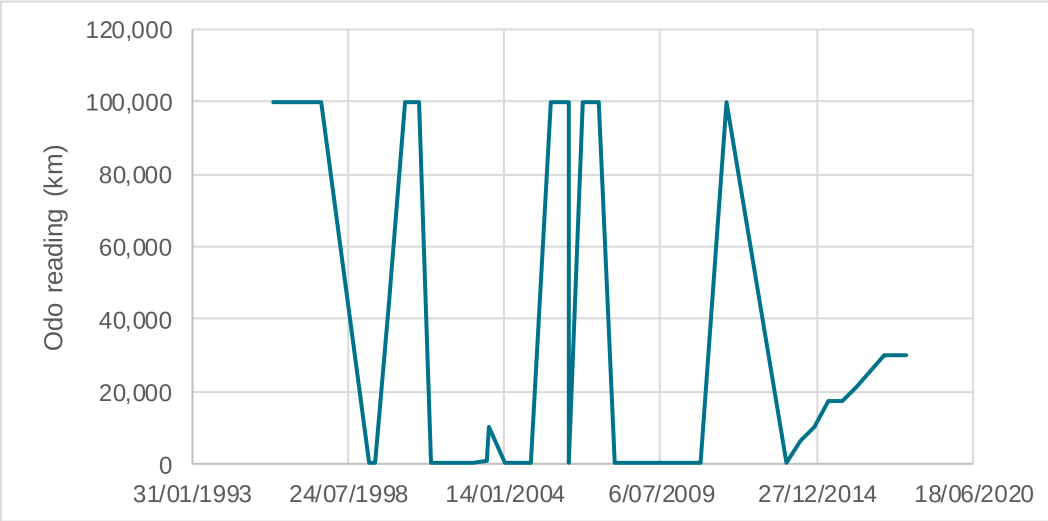


These falling readings were projected forward from the previous valid record, and all following records which still showed a positive trend were shifted to start from the new projected endpoint.

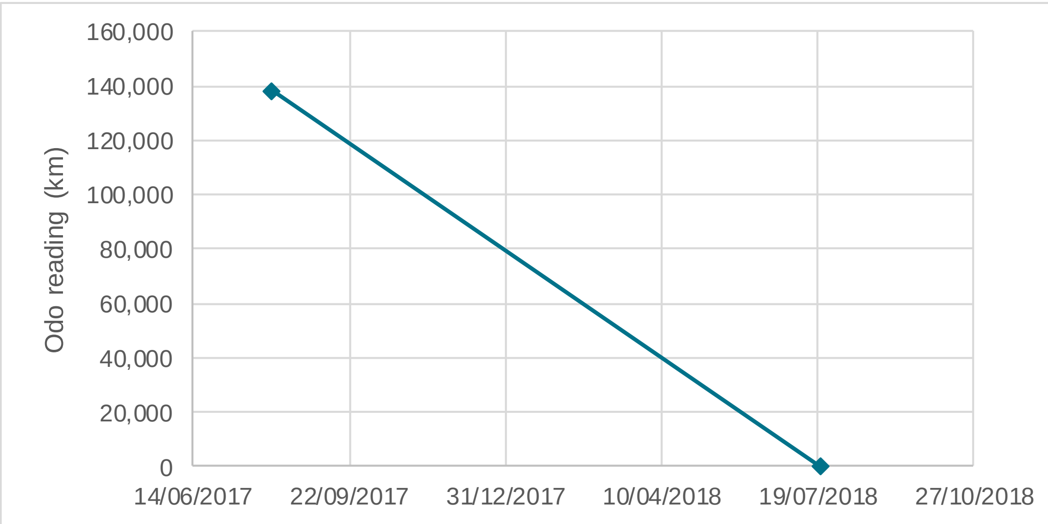


B.5 Unresolved errors

Some records could not be resolved due to their erratic nature, or high complexity due to time constraints. A few examples include:



No discernible trend.



A lack of data points.

Of the 109 million records, there are approximately 491,000 unresolved irregularities relating to 352,000 vehicles.

B.6 Projected readings

In a number of cases it was necessary to project odometer readings where these were not recorded – for example where a vehicle was deregistered and a final odometer reading at deregistration was not recorded, or where a vehicle is new and few warrant of fitness checks have been undertaken. These projections were based on the typical mileage for a vehicle of that age and/or the actual experience to date for that particular vehicle.

C Potential improvements

In this paper we have attempted to undertake a multivariate analysis i.e. to model the impacts of a number of factors simultaneously e.g. vehicle year/age, UCSR and vehicle mass. However, it is not an ideal multivariate analysis in that it only focuses on vehicle characteristics, whereas in reality there are many other factors which are likely to be predicative of injury and fatality numbers.

Two simple but major improvements to the data supplied for this analysis could greatly improve the modelling outcomes:

- Increasing the number of fields provided in the MVR and CAS extracts.
- Providing a unique link (e.g. a license plate) between the MVR and CAS datasets.

Potential improvements with additional fields are discussed in Section C.1 below. Providing a link between the MVR and CAS data would greatly improve the quality of the vehicle data stored in CAS. For example, where a vehicle in CAS is simply specified as a 'Honda' with no stated model type, knowing the license plate would enable the specific model to be looked up in the MVR. This would mean that the proportions of vehicles by various categories (e.g. categorised by UCSR) would be more consistent between the exposure and accidents observations.

We recognise that there are privacy considerations here in regard to license plate details, although this could be addressed by providing scrambled or anonymised license plate numbers (so long as they have been scrambled in the same way in both the MVR and CAS datasets).

C.1 Additional fields

Some additional fields from the MVR which would be useful are:

- Vehicle mass
- CC rating
- Vehicle usage
- Deregistration date
- Registered address details (at a high level e.g. suburb or post code)
- Registered owner details (age/sex)

Some of these fields were estimated by mapping the vehicle make/model combinations between the odometer data and the NZ open fleet data, although for some make/model combinations these mappings were somewhat ambiguous.

Some additional fields from CAS which would be useful are:

- Details of the injured/fatal occupants (age/sex/license/etc)
- Crash date and time
- Crash location
- Road details (intersection/speed limit/condition, etc)
- Presence of alcohol
- Factors contributing to crash e.g. speed