Estimation of the crash rates of vehicles considering vehicle age, crash period and vehicle cohort

Anderson, R.W.G.

Centre for Automotive Safety Research, The University of Adelaide, South Australia, 5005

Abstract

What will the future crash rate be of a vehicle built today? Or one built 5 years ago, or in 3 years time? Such questions are relevant to the costs and benefits of new safety technology, and to understanding future crash risks more generally. They are also a key to distilling the effect of vehicle safety from other factors in road safety. Crash rates have a few important dimensions; when viewed along these dimensions, the underlying patterns in the rates provide a means of estimating present and future crash rates of vehicles. These dimensions are:

- the age of the vehicle, because it relates to factors including vehicle kilometres travelled, typical driver age etc.,
- the period (calendar year) being considered, because it relates to factors including prevailing levels of compliance, changes to infrastructure, the characteristics of other vehicles in the fleet, and
- the cohort (year of manufacture) to which the vehicle belongs, as it relates to levels of crashworthiness and primary safety.

Analogous age-period-cohort data are also used in public health research to study trends in the rates of disease. In this paper, age-period-cohort vehicle data are assembled from mass crash data from NSW. The data are used to impute crash rates by cohort and by period. Rates are imputed for passenger vehicles built between 1999 and 2009 and by projection to 2020. The results suggest that large reductions in crash rates may be expected for future cohorts of vehicles.

Introduction

The change in the number of crashes from year to year is subject to systematic changes in crash risk (along with random effects). These changes in risk arise from changes to a road transport system that includes the characteristics of the roads, the vehicles that drive on those roads, the drivers that operate the vehicles (and other road users), and the regulations and enforcement that govern the system. Each element is in constant interaction with the other elements of the system.

Vehicles are something of a special case within this system as there are lags between improvements to new vehicles and the benefits of those improvements. A change in speed limit or enforcement may act with almost immediate effect and may affect all vehicle users, but an improvement to new vehicles will have benefits spread out over many years, benefiting only new vehicle drivers at first, with the benefits accruing over many years after the vehicle is first sold. During the period in which the vehicle is in service, other elements of the system will also change. Hence the crash experience of a particular vehicle at any given time is influenced by its inherent safety and the evolving system in which it operates. Examining crash rates of all vehicles in a given period (i.e. a cross-sectional study of crash rates) can provide information the average safety of all vehicles in operation during that period, but may mask the heterogeneity of risks across vehicles, and furthermore is unlikely to be a useful guide to the crash experience of new vehicles whose service life lies in the future. Hence, if historical crash data are to be usefully used to answer questions regarding the likely number and timing of crashes involving vehicles built in the future, such data need careful treatment.

Crash involvement of groups of vehicles are of interest for several reasons. Whether or not to regulate a particular vehicle technology may rest on benefits outweighing the costs of the technology, and trends in cross-sectional data are often used in support of such regulation (e.g. Fitzharris and Stephan, 2012). Similar questions are pertinent in the area of motor vehicle insurance; insurance actuaries may have an interest in knowing the probable number and timing of crashes involving vehicles of particular build years and ages. Such information might aid in the risk ratings applied to those vehicles. However, available studies show that the effect of build year and age often appear to be conflated in vehicle insurance risk models (Kahane et al., 2007; Yeo et al., 2001).

Ideally, crash reductions expected from a technology would apply to an estimate of crashes thought likely to occur to vehicles in the future, rather than how the technology might have affected crashes in the past. Therefore both the number and timing of crashes occurring amongst future cohorts of vehicles will affect discounted crash cost calculations. If crash rates are in decline due to changes to the system, and in addition, if the crash rates of vehicles are improving with build year, then using historical data to estimate the crash burden faced by vehicles to be built in the future may be result in inflation of future risks, and double-counting of benefits.

Useful methods for dealing with these kinds of problems have been developed in the area of public health. Rates of human disease amongst a particular cohort in the population often depend on characteristics of a particular cohort in question (represented by year of birth), the age of the cohort (years of life elapsed) and factors associates with a given period (e.g. the current calendar year). When rates of disease depend on these three factors, it useful to use Age-Period-Cohort (APC) techniques to describe incidence data. Several publications can be found on the implementation of such techniques (e.g. Carstensen 2007; Palmore, 1978).

There is a direct analogy between the factors considered in an APC analysis of disease with vehicle effects in crashes:

- The age of a cohort of vehicles might affect their crash rate in a given period (from factors including annual vehicle usage, type of driving and average driver characteristics that might vary with vehicle age),
- The period being considered will affect the crash rate (from factors including prevailing average speeds, levels of compliance, changes to roads, the nature of other vehicles and road users), and
- Characteristics of the cohort of vehicles themselves will affect their crash rate (from factors related to the design and performance of the vehicle, particularly the improved crashworthiness of vehicles, and crash avoidance features).

Note that the age of the vehicle is simply the difference between the crash year (period) and the year that the vehicle was manufactured (cohort).

Aim

The objective of this paper is to introduce the concept of APC modelling to the analysis of crash data and to illustrate the characteristics of crash data when disaggregated according to vehicle age, vehicle cohort and period. Further, the resulting tabulations are used to look at trends in lifetime crash involvement of various vehicle cohorts.

Data and Methods

The present analysis uses eleven years of crash data (1999-2009; injury and fatal crashes) from New South Wales. APC analysis involves the disaggregation of data on the incidence of disease (or of crashes) according to age, period and cohort. If incidence data for each age group are tabulated against the period, then the incidence data for each cohort as it ages are contained on the corresponding diagonal. This tabulation is known as a Lexis diagram (See Figure 1). Once the data have been correctly tabulated along these lines, the table can be rearranged to tabulate incidence data for each cohort over period, or for each cohort over age.

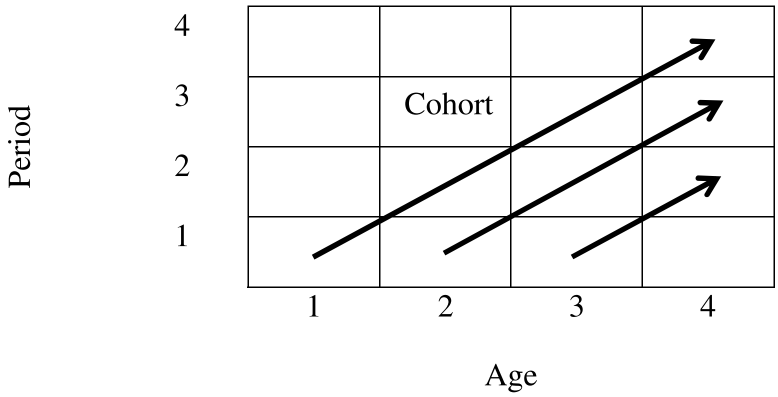

Figure 1. Layout of a Lexis diagram

Results of APC disaggregation on NSW crash data

Figure 2 shows the results of such a disaggregation and consequent tabulations of the NSW crash data on injury and fatal crashes from P 1999-P 2009[1]. Six plots of crash rates are shown:

- by period (calendar year) for various ages,
- by cohort for various ages,
- by age for various periods,
- by age for various cohorts,
- by period for various cohorts, and
- by cohort for various periods

Note that age zero in this paper refers to the first year of life, and various adjustments have been made to account for the partial year that most vehicles in the cohort would have experienced. Furthermore, the number of vehicles in the cohort at the end of year-one is used to define the size of the cohort (see later discussion).

These graphs are an unusual but revealing way of examining crash data. Note that as each plot is between two of the three factors (age, period and cohort), but each need to be interpreted with the influence of the third factor in mind.

[1] In this paper, the prefix P will be used when referring to a period, and C will be used when referring to a cohort.

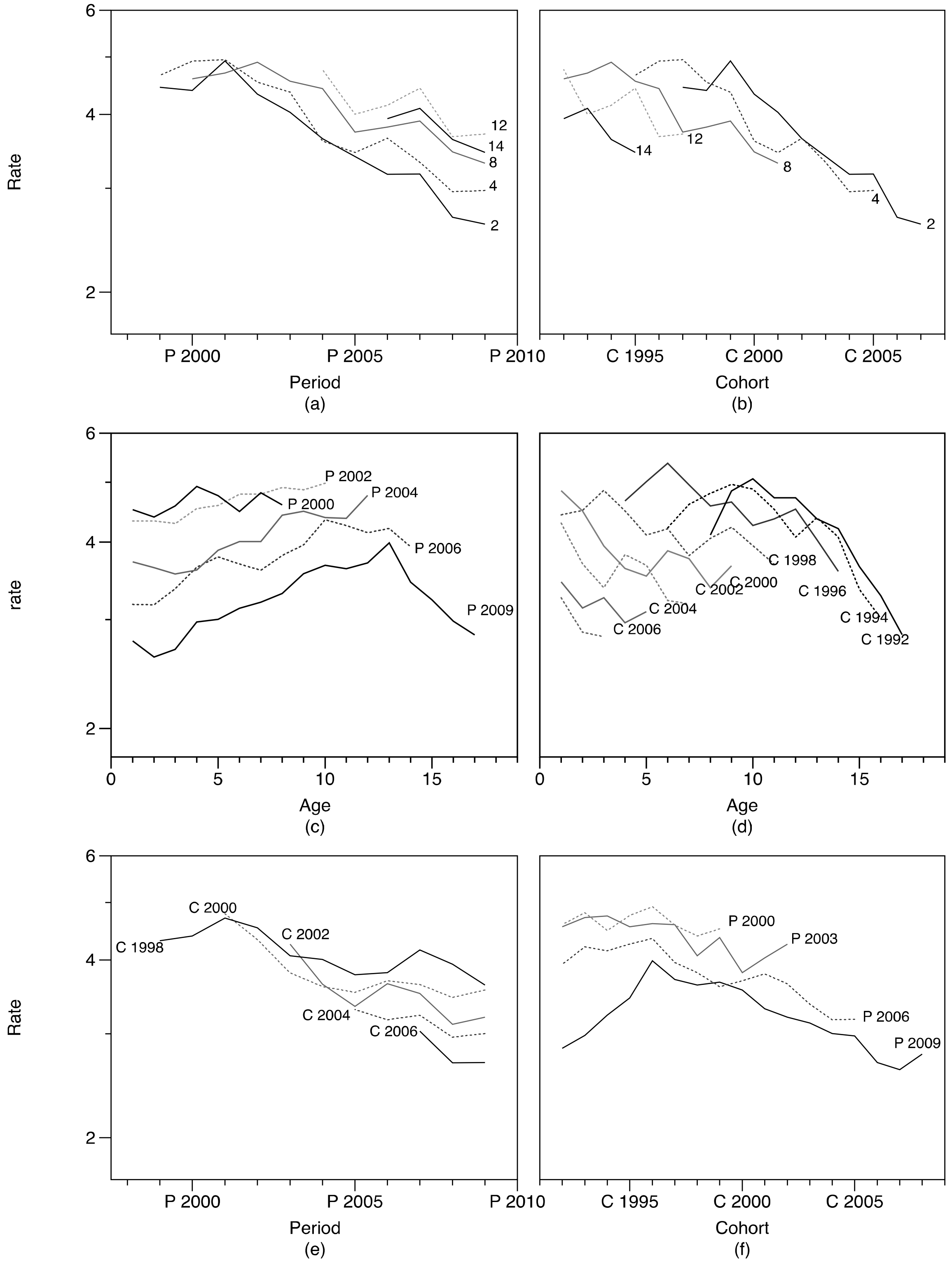

Figure 2. Crash rates (injury and fatal) of passenger vehicles (left) by period for different vehicle ages and (right) by vehicle cohort for different vehicle ages

Some comments on Figure 2 are as follows:

- Referring to (a): crash rates of vehicles in every age group have declined with period (noting that (b) at each vehicle age, crash rates are generally lower for successive cohorts). This points to a strong effect of period and cohort on the crash rate. An age effect is apparent too with the peak crash rate in each period occurring at around 12 years of age. However, for each cohort, the crash rate appears to decline with increasing age (b).

- The peak in the crash rate is also discernable in plot (c). Since about P 2002, the crash rate at every age has declined. (This is another expression of the effect of period and cohort.)

- The local peak in the crash rate in P 2002 is evident in (d) with Cohorts C 1992 - C 1998 exhibiting peaks in their respective crash rates at ages that correspond to P 2002. Age specific rates for Cohorts from C 1998 onward show lower crash rates at each age (again showing the effect of age and period in the data).

- Crash rates are declining in each cohort with period (and age); (e). In recent periods the higher crash rates in older cohorts is apparent. In (f), the decline in crash rate with cohorts younger than 12 years of age is clearly apparent in every period.

By inspection of these figures, it is apparent that all three factors – age, period and cohort – are likely to be influencing crash rates. If that is so, then APC analysis of crash data faces a hurdle: as age is only the difference between period and cohort, and because all three factors may influence crash rate, APC analysis can isolate the effect of two factors only when the size of the effect of the third factor is specified prior to the analysis.

Calculation of the number and timing of crashes over the life of a cohort of vehicles

Fortunately, for various types of analysis, it may not always necessary to disentangle all three effects. For example, one of the objectives of applying an APC analysis on crash data could be to examine crashes over the service lives of cohorts of vehicles. With sufficient data, the age-specific rates can be plotted for an individual cohort. Such a plot would be the result of the aging of the cohort and the effect of period, as the rate at each additional year in the age of the cohort would also represent the effect of changes attributable to an advance in the period by one year. In this case the effect of period may be allowed to remain bound together with age, and some projection of trends can be made as long as there is sufficiently consistency in several aspects of the effects of age and period within each cohort and over time.

However, a complication can arise because, as is clear in Figure 2, often only fragments of some cohorts' crash histories will be available in crash data. Particularly, cohorts of vehicles built in recent periods have not yet completed their service lives and hence their lifetime crash involvement is not yet known. For example, the data used in this paper contain the crash history for the first 11 years of C 1999, the first three years of C 2007 and years 9 to 20 of C 1990.

A feature of the pattern of the crash involvement of cohorts of vehicles is that, over succeeding cohorts, it appears quite stable: the shape of the age distribution of crashes for each cohort appears similar, and only the scale appears to change. (This phenomenon will be illustrated in the Results.) If the stability in the shape of the age distribution is assumed, it allows future crash involvement to be estimated from historical crash records. Hence averaging the shape of known fragments of age distribution of crashes of all cohorts can be used to build an estimate of the crash history profiles for any given cohort, based only on a partial history of the cohort.

To assume that the shape of the distribution of the age-related crash rate of a cohort is stable implies two further underlying assumptions to be made: That the underlying age effect is stable between cohorts and that the period effect is constant and linear (for example, a constant per-period decline in crash rate). It is posited here that both assumptions are necessary to produce a stable distribution of the age-related crash rate of all cohorts.

In order to estimate the lifetime crash involvement of various cohorts, the following steps were made:

- First, the crash rates in each year of age for each vehicle cohort were calculated for the crash data that were available,

- Second, the *proportional changes* in the crash rates from one age to the next in each cohort were examined, and then

- these were averaged to derive an crash distribution of a single cohort over its age.

Examining the proportional change in the crash rate is, in effect, a method for defining the shape of the distribution.

Determining the size of the cohort (the denominator in the rate)

Of particular importance to assessing vehicle technology is the number of relevant crashes per vehicle in the cohort (the crash rate). Hence, in this study, the number of vehicles entering the cohort in NSW is the relevant denominator when calculating the per-vehicle crash rate.

The numbers of vehicles in each cohort for several calendar years are available from the ABS vehicle census data. However, the total number is not precisely known as the number in the cohort builds over 1-2 years as they appear on State vehicle registers, and during this period, some vehicles will be transferred interstate and others may cease service due to crashing or failing in some other way.

Notwithstanding the above, for the purposes of this study, the size of the cohort in the census taken at the end of the cohort's second year of life was used to represent the vehicle cohort population, as it is at this age that the number of vehicles peak in each cohort.

Adjusting crashes to account for interstate transfers

Another complication can arise because of the choice of the denominator in the rate calculation. If the analysis is about the crash incidence of *all* vehicles built in a particular year, then the choice of denominator will be the total number of vehicles in the cohort. NSW is a net exporter of second hand vehicles. Hence vehicles in cohorts originating in NSW may crash in a jurisdiction other than NSW. Put another way, the crash experience of vehicles within one cohort originating in NSW may be underestimated if only crashes within the state are used.

Adjustments were made to NSW crash numbers to reflect the probable effect of the movement of vehicles out of NSW. The details of the method are not included here, but in summary:

- ABS motor vehicle census data were used to construct survival functions of registered vehicles in NSW and in Australia overall.

- It was assumed that the true rate of attrition of vehicles sold originally in NSW is represented by the survival function of vehicles for Australia as a whole.

- The difference between the apparent survival function of vehicles in NSW and the survival function of vehicles in Australia represents the movement of vehicles from NSW to other States. (This difference is the "export function".)

- It was assumed that these exported vehicles crash at similar rates to those remaining in NSW, and the export function was used to adjust numbers of crashes recorded in NSW to represent crashes, in all states, of vehicles originating in NSW.

Results of estimating crash rate distributions and the lifetime crash rate of several vehicle cohorts

After selecting the appropriate denominator and adjusting crash numbers to reflect interstate transfers, crash rates for each cohort were calculated for the period 1999-2009. Based on this, the average crash involvement of all cohorts was calculated and assembled to create an average distribution of crashes over vehicle cohorts' first 25 years of life.

The resulting distribution is shown in Figure 3.

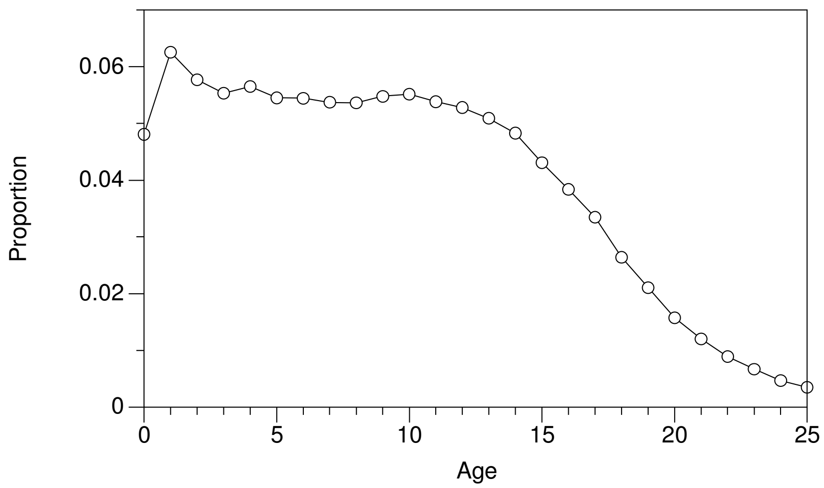

Figure 3. The crash distribution (injury and fatal) over age of a vehicle cohort in New South Wales based on the average year-on-year changes in crash numbers across vehicle cohorts in crash years 1999-2009. Proportion expressed is of crashes within the first 25 years of life.

The individual fragments of the crash rate history available in the period P 1999 – P 2009 are shown in Figure 4. These figures suggest that the average distribution may slightly misrepresent any single cohort's crash profile. Earlier cohorts appear to have more crashes than the average in early years. By implication, later cohorts may have a greater proportion of their total number of crashes later in their service lives, consistent with an increasing service life. However, for the present purpose the average distribution will be used.

Trends in overall crash rates in successive cohorts of vehicles

Having estimated a distribution the average crash involvement all fragments, the remaining crash history (outside of the fragment) of any given cohort was estimated. An illustration is given below.

For the cohort of vehicles with a year of manufacture equal to C 2003, there were seven years of available crash data (P 2003 to P 2009). The total crash rate over those seven years was 22.7 crashes per 1000 vehicles in the cohort. The average injury crash distribution can be scaled such that the cumulative crash rate at 7 years is also 22.7 crashes per 1000 vehicles. The scaled average distribution for year eight to year 25 may then be used to estimate the future crash experience of the 2003 vehicle cohort.

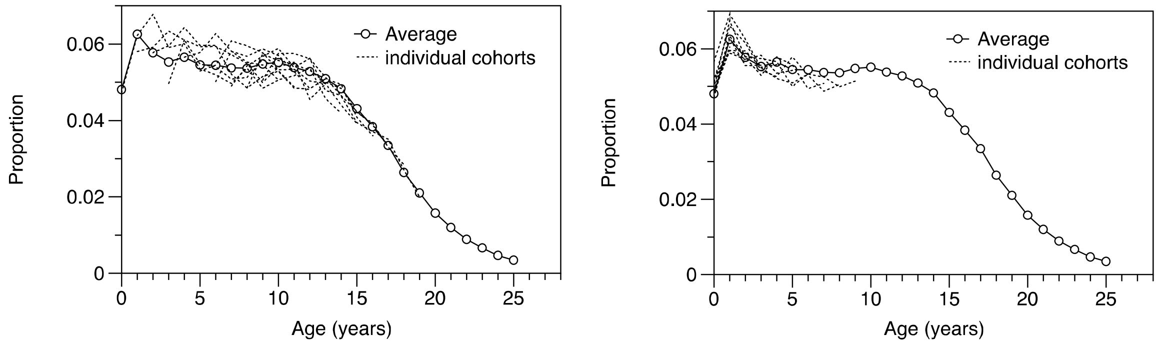

Figure 4. The crash rates of individual cohorts overlaid on the average curve: for cohorts C 1990 to C 1999 (left) and for cohorts C 2000 to C 2008 (right)

The crash distribution of the C 2003 cohort, the fitted average distribution and the estimated distribution are illustrated in Figure 5.

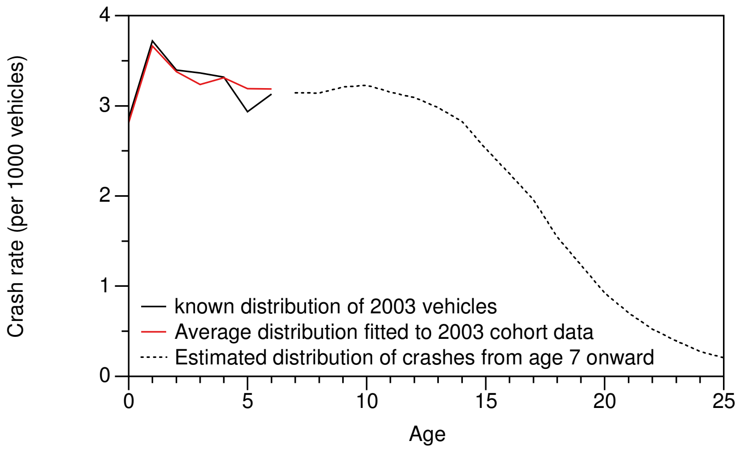

Figure 5. The estimated distribution (injury and fatal) of crashes amongst the vehicle cohort C2003 in New South Wales. The first seven years is shown, with the relevant portion of the average distribution scaled to fit. Based on this fit, an estimation of the future distribution of crashes within the cohort has also been drawn.

Finally, the total number of crashes, or the overall crash rate can be imputed. In the above example, the total number of injury crashes likely in the C 2003 cohort (over the first 25 years of the cohort's life) is 58.5 per 1000 vehicles.

Lifetime crash involvement

This procedure was repeated for all cohorts with a year of manufacture falling between C 1999 and C 2009, and the imputed number of crashes likely in the first 25 years was calculated. The result for each cohort and the trend in the total are shown in Figure 7. Similar results were obtained for fatal crashes. Note that the estimated crash rate for the C 2009 cohort has been omitted as it is based on a very small number of crashes. A projection of the 25-year rate is shown for cohorts to C 2020.

It is worth noting again that the declining crash rate over cohorts is a combination of cohort effects (attributable to year of manufacture; i.e. inherent primary and secondary safety) and period effects (attributable to the period in which crashes within a given cohort occur) to the end of P 2009.

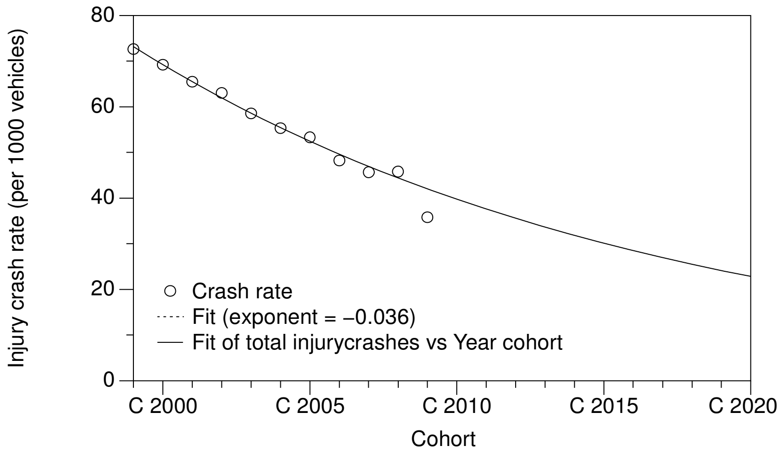

Figure 6. Crash rate (injury and fatal) of vehicle cohorts over the first 25 years of life, estimated for vehicles built from C 1999 to C 2009 and a projection to C 2020 (C 2009 has been omitted from the trend calculation)

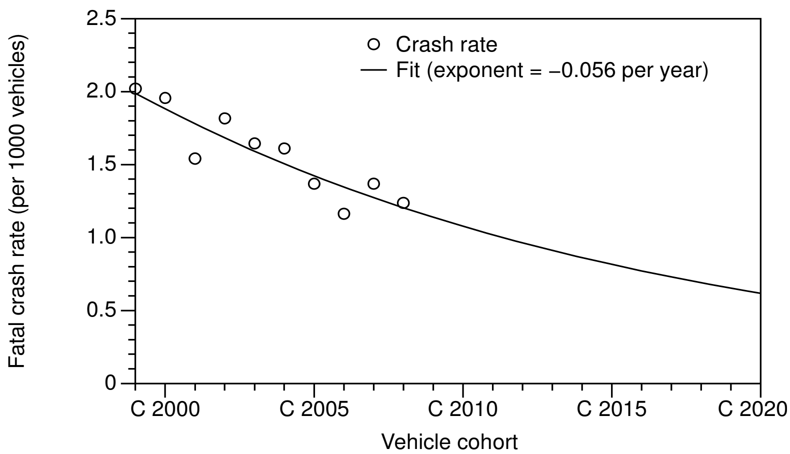

Figure 7. Fatal crash rate of vehicle cohorts over the first 25 years of life, estimated for vehicles built from C 1999 to C 2008 and a projection to C 2020

Discussion

The purpose of this paper was to introduce age-period-cohort techniques to the area of vehicle crash rate analysis. The paper serves only as an introduction to the possibilities of such an approach to the area of road and vehicle safety research. It is proposed that such analyses are more appropriate when attempting to estimate of the costs of injury in a way that is relevant to assessing the benefits of new technology, than the use of secular trends in cross-sectional (i.e. across cohort) crash data. As has been shown, there is heterogeneity in crash rates across cohorts in any given period, and hence cross-sectional data are likely to be an inappropriate baseline for calculating benefits of mitigation technologies only available in future cohorts of vehicles.

By disaggregating crash data using the principals of APC analysis, this paper has shown trends in lifetime crash rates of passenger vehicles in NSW. The trends in the lifetime crash rate of vehicle cohorts is such that is appears that vehicles sold new in 2009 will experience about 43% fewer injury and fatal crashes per vehicle than the C 1999 cohort. The declines are likely to be a product of both the effect of vehicle design changes and the ten-year difference in the evolution of the road system, which includes effects of changes to roads, other vehicles, behaviour changes, enforcement and also changes in per-vehicle driving exposure. Projections of the trends suggest further substantial declines in crash rates for future cohorts of vehicles.

The methodology used to estimate the lifetime crash rates of vehicles in this paper can be applied to a range of analyses. Rates of crashes of particular types involving particular vehicle classes could be examined, and/or rates of crashes in particular environments. These may be of particular importance to specific vehicle technologies; for example, forward collision technologies designed to operate in urban environments. However, further study is likely needed to refine the methodology presented above:

- The assumption of a constant age-period effect may be questionable. Longer vehicle lives due to improving reliability may mean that exposure in later years of life may increase with cohort year. Discounting would mitigate such an effect to some extent, but the effect in the present analysis would be to increase the lifetime crash rate, particularly of later cohorts.

- A constant period effect is likewise open to question, and needs examination.

Note that such assumptions are currently made implicitly in projecting cross-sectional (i.e. all cohorts in a given period) crash rates into the future. An advantage of APC analysis is that such assumptions can be examined systematically.

Conclusions

Age-Period-Cohort analysis provides a method of disaggregating crash data to reveal patterns and trends. It is, in theory, a superior approach to estimating the crash rates of vehicles built in particular years, and hence for estimating the benefits of additional technologies designed to reduce the risk of injury in specific crash types.

Acknowledgements

This work is based on a prior study of the benefits of forward collision avoidance technologies performed under a contract from Transport and Main Roads, Queensland. Further development has been made possible through funding by the South Australian Department of Planning, Transport and Infrastructure and the Motor Accident Commission of South Australia. The views expressed in this paper are mine, and do not necessarily reflect the views of the abovementioned organisations. I am grateful to Dr Paul Hutchinson for his thoughtful advice during the preparation of this paper.

References

Carstensen, B. (2007). Age–period–cohort models for the Lexis diagram. *Statistics in medicine*, *26*(15), 3018-3045.

Fitzharris, M., & Stephan, M. K. (2012). Assessment of the need for, and the likely benefits of, enhanced side impact protection in the form of a Pole Side Impact Global Technical Regulation.

Kahane, Y., Levin, N., Meiri, R., & Zahavi, J. (2007). Applying Data Mining Technology for Insurance Rate Making: An Example of Automobile Insurance. *Asia-Pacific Journal of Risk and Insurance*, *2*(1).

Palmore, E. (1978). When can age, period, and cohort be separated?. *Social Forces*, *57*(1), 282-295.

Yeo, A. C., Smith, K. A., Willis, R. J., & Brooks, M. (2001). Clustering technique for risk classification and prediction of claim costs in the automobile insurance industry. *Intelligent Systems in Accounting, Finance and Management*, *10*(1), 39-50.