Assessment of the ICBC Report Regarding the Safety of Right Hand Drive Vehicles in BC

Mohua Podder, PhD and Rick White, M.Sc.
Statistical Consulting and Research Laboratory
Department of Statistics
University of British Columbia
6356 Agricultural Road, Vancouver, BC

March 26, 2010

1 Introduction

This report evaluates the analyses done by ICBC in their report titled "The Safety of Right-Hand-Drive Vehicles in British Columbia". The request for the evaluation came from the Imported Vehicle Owners Association of Canada (IVOAC). The number of right-hand-drive (RHD) vehicles being imported into Canada over the past few years has been increasing. Imported vehicles over 15 years old are exempt from the requirement of having a manufacturers plate or letter stating that the vehicle is compliant with the Canadian motor vehicle safety standards. This allows RHD vehicles to retain their RHD control configuration. The belief is that potential issues related to these vehicles might lead to a greater accident risk for operators of RHD vehicles and people in RHD vehicles might be more severely injured than people in LHD vehicles.

ICBC conducted a study in order to assess the safety issues of RHD vehicles in BC and published a report in early 2008. The main purpose of the study was to compare RHD to LHD vehicles in terms of their risk of crash involvement and their occupant protection potential. The study applied three different methodologies to try to answer these questions: a Relative Risk Analysis comparing RHD and LHD culpable crash rates for the same group of drivers; a Survival Analysis to compare the time to first culpable crash between drivers of RHD and LHD vehicles; and a Poisson Regression Analysis to compare the risk of RHD vehicles to a similar group of LHD vehicles and their drivers.

2 Data Issues

The data for all three analyses were obtained from the ICBC crash-claim data. Does this database accurately reflect all crashes in BC or just those that are reported? Drivers may choose not to report their crash to avoid an increase in insurance rates because the value of the vehicle is low or the repair is inexpensive relative to the increased insurance cost, perhaps the driver doesn't have the appropriate insurance coverage. If the tendency to not report a crash is related to the drive configuration of the vehicle, it may introduce a bias in the data that could be reflected in the analyses.

All three analyses rely on the definition of culpability. A vehicle is culpable in a crash if it is assigned at least 50% of the responsibility during the claim adjustment process. Is blame assigned by objective criteria or is there a subjective component? If the assignment of responsibility is affected by the drive configuration of the cars involved in the crash then all analyses based on this definition will be inherently biased. The assignment must be based on the circumstances of the crash only. That means if a LHD car involved in a crash is deemed non-culpable then that status would not change if that car had been a RHD vehicle.

3 Relative Risk Analysis

The first method used looks at the relative risk of culpable crashes to non-culpable crashes between RHD and LHD vehicles operated by the same driver. This method compares crashes within drivers to control for driver differences. All drivers involved in crashes while operating a RHD vehicle since 2001 were included in the analysis. In addition the principal operators (POs) of these RHD vehicles, if not already included, were selected to provide a set of drivers not involved in a crash while driving a RHD vehicle. The complete crash history of each driver since 2001 was used in the analysis. Each crash was classified as culpable or non-culpable from the target driver's perspective.

After applying several constraints to the data, 359 crashes for RHD and 1204 crashes for LHD vehicles were identified. Of the 1204 LHD vehicle crashes, 324 were from drivers with RHD vehicle crashes and 880 were from the RHD PO group, the drivers without a crash while driving a RHD vehicle. The total number of drivers in each group is not mentioned. A cross tabulation of culpable/non-culpable crashes versus RHD/LHD vehicle was created. A Chi-squared test of independence was used to determine if the type of crash was associated with the drive configuration of the vehicle. This analysis assumes that all events in the table are independent events not connected by any other factor whereas here we actually have several 2x2 tables, one for each driver. The data within each table depends on a specific driver but the tables themselves are independent. In essence, we have a stratified sample. In order to perform the Pearson's Chi-squared test for independence, the stratification is ignored and the individual 2x2 tables are summed across drivers to make a single 2x2 table. A more appropriate analysis for this type of data would account for the stratification within the data. A possible analysis method would be the Cochran-Mantel-Haenszel test.

Another possible concern for this data is if two vehicles involved in the same accident are included in the data. This is probably a rare event and therefore a minor concern but should probably be checked. The 2x2 tables for any drivers involved in the same accident would no longer be independent and the analysis would need to be adjusted accordingly.

4 Survival Analysis

The second method of analysis uses a Cox proportional hazard regression model to compare the instantaneous crash rate of RHD and LHD vehicles after they are first insured. The response data is the time to a culpable crash following the initial insurance policy purchase by a PO for each vehicle. The model has the advantage of including all vehicles even those that never had a culpable crash over the course of the study. In addition, the model can also include adjustments for other factors that might modulate the effect. In this analysis data were collected from "all RHD POs aged 20 years and older at the time of first policy and all vehicles (RHD and LHD) for which they were listed as POs". Only culpable crashes were included as an event. Age and gender were included as adjustments in the final model.

The sampling method described in section 2.2 sounds like a small set of POs was selected for the analysis, each having been a PO of several vehicles of which at least one was a RHD vehicle. The description of the analysis in section 3.2 implies a different set of drivers was included. Section 3.2 claims "A total of 23717 drivers were included in the analysis of which 2882 were associated with RHD vehicles". This section also describes the results in terms of vehicles. It is not clear who has been sampled or what is being modeled here. Is it time to first culpable crash of a driver or vehicle? Are multiple vehicles of the same driver or PO included in the model? Without a clear description of the sampling scheme or data being collected, it is hard to comment on the methods.

If the data is time to first culpable crash of a vehicle and several vehicles are associated with the same PO, a correlation structure is induced into the data which needs to be adjusted for in the model. Another concern with POs having multiple vehicles is they may split their time at risk between their vehicles which further complicates the analysis. If we think of POs as a stratification variable, a stratified Cox model can be applied to deal with some of these issues.

Ignoring the stratification issue, the analysis assumes that each vehicle is at risk at all times when in fact a vehicle is only at risk when it is being driven. A vehicle driven 5 days a week is more likely to have a crash than one driven once a week. The crash risk of a vehicle is also affected by where and when it is driven. A vehicle driven in rush hour on Monday morning probably has a different risk than one driven on Sunday morning, as would a vehicle driven in downtown Vancouver compared to one driven in White Rock. Is there a way to guarantee that RHD vehicles are driven the same amount of time, under the same conditions and locations as LHD vehicles? If not, what is the affect of these factors on the results?

Age and Gender were included in the model as covariates. It is unclear if the effects presented in Table 3 are from a model that contains only main effects or one that contains interactions with RHD vehicles. As main effects, they modulate both the LHD and RHD crash rates. So all statements pertaining to the effect of age and gender on RHD vehicles apply equally to LHD vehicles. A model with interaction allows separate age and gender effects to be estimated and compared for the RHD and LHD vehicles. The report states that the RHD vehicle group contained more males and was younger. This causes partial confounding between age, gender and RHD vehicles. Confounding affects the model estimates especially in a model with only main effects. The model estimates the crash rate to be higher for males and younger drivers. Confounding may cause the model to attribute some of these effects to RHD vehicles.

5 Poisson Regression Analysis

In the final analysis, a Poisson Regression model is used to compare the crash rates between RHD vehicles and a group of similar LHD vehicles. A RHD group consisting of 748 vehicles, was identified and a LHD group consisting of 8933 vehicles, was selected by model, make, year and body style so the proportion of vehicles in that group matched the proportions in the RHD group. The response data is the number of crash-claims in a two year period following the date of a PO's first policy with that vehicle. Each crash was classified as injury or material damage only and culpable or non-culpable. Covariates considered in the model were gender, age, region (lower mainland or not), speeding contraventions and non-speeding contraventions.

The main concern of the model is to compare the crash rate between RHD and LHD vehicles. However matching model-year-style in vehicles does not guarantee that both RHD and LHD vehicles are exposed to equal amount of driving time or are driven in the same locations or in the same traffic conditions. The confounding affect of gender and age with RHD vehicle operation also affects a Poisson regression model in a similar fashion as a Cox model. The distribution and effects for region and traffic contraventions is not presented in the report.

Another concern with Poisson regression models is overdispersion in the response data. If the counts are overdispersed than an overdispersed Poisson or Negative Binomial regression model needs to be fitted otherwise the standard errors of the effects will be underestimated and effect significance will be overstated. The report does not indicate if overdispersion was checked in the model.

The analysis is done by vehicle not by driver. The Poisson regression model assumes the data for each vehicle is independent from each other. However a vehicle can have many drivers and a driver can operate many vehicles. This may introduce a correlation between the vehicles if the same driver operated more than one vehicle in the study. If there are many such drivers, the induced correlation could become an issue for this model as well.

The report says principal operators were also examined to investigate the differences between RHD and LHD vehicles at the driver level but it doesn't explain how this is done. Does a RHD driver always operate a RHD vehicle? If not, how is a crash in a LHD vehicle by a RHD driver dealt with in the analysis?

6 Summary

There are several issues that could affect the results of the study conducted by ICBC in the report "The Safety of Right-Hand-Drive Vehicles in British Columbia". It is possible the conclusion would remain unchanged even after these issues were resolved but we do not know without actually doing those analyses.

Reporting a crash to ICBC is not mandatory if an insurance claim is not made. Are the factors that affect the tendency to make a claim related to the drive configuration of the vehicle? How is blame assigned in a crash claim? If blame assignment has a subjective component, is it related to drive configuration of the vehicle? If either is related to drive configuration then the data contains a bias that could affect the results of any analysis based on the data and may not reflect the true risk of a culpable crash of LHD or RHD vehicles in the province as a whole.

In addition to data issues, there are some issues with the analyses themselves. The relative risk analysis completely ignores the repeated measures within each subject. This analysis is easily corrected by using a Cochran-Mantel-Haenszel test instead of a Chi-squared test. There is also a minor issue of some data coming from the same crash but this is probably a rare event and will have little impact on the results.

The problems with the other two analyses are more subtle and technical. The drive configuration of the vehicle was not evenly distributed between gender and age. Although the model adjusted for these factors, imbalance can still affect the estimates in the model. If the data had been matched on age and gender, the results would be more trustworthy. A stratified or paired analysis might then be more appropriate in these cases. A More subtle issue is risk exposure. A vehicle is only at risk of a culpable crash when it is driven. The risk exposure also depends on driving locations and conditions. A vehicle driven to work every day during rush hour is at greater risk than one driven once a week on Sunday morning. It is unclear if adequate data can be obtained to adjust for these factors but they could have a big impact on the results. Drivers using multiple vehicles or vehicles driven by multiple drivers is another issue. This introduce a correlation into the data that needs to be taken into account. If there are several such vehicles or drivers in the data and the repeated measures are properly adjusted for, the results of the analyses could be quite different. Again it is unclear if such data is available.

Overall, the ICBC report suggests that RHD vehicles and their operators are at a greater risk than their LHD counterparts but issues with the data and the analyses suggest that further study is needed. Causation is difficult to establish with observational data. I would caution the use of the ICBC report as anything more than an indication that further study is needed.