TRAFFIC INJURY RESEARCH FOUNDATION

NOVA SCOTIA ALCOHOL IGNITION INTERLOCK PROGRAM:

OUTCOME EVALUATION TECHNICAL REPORT

TIRF

The knowledge source for safe driving

The Traffic Injury Research Foundation

The mission of the Traffic Injury Research Foundation (TIRF) is to reduce traffic-related deaths and injuries. TIRF is an independent, charitable road safety institute. Since its inception in 1964, TIRF has become internationally recognized for its accomplishments in identifying the causes of road crashes and developing program and policies to address them effectively.

Prepared by:

Ward Vanlaar, Marisela Mainegra Hing, Robyn Robertson

Traffic Injury Research Foundation
171 Nepean St. Suite 200
Ottawa, ON K2P 0B4
Ph: (613)238-5235
Fax: (613)238-5292
Email: tirf@tirf.ca

www.tirf.ca

December 2014
Traffic Injury Research Foundation

ISBN: 978-1-926857-66-4

EXECUTIVE SUMMARY

Introduction

In September 2008 Nova Scotia's Alcohol Ignition Interlock Program was implemented. The overall objective of the Program was to improve road safety and reduce the number of road traffic crashes and fatalities that may occur due to impaired driving. This report describes the outcome evaluation of Nova Scotia's interlock program. The Traffic Injury Research Foundation (TIRF) conducted this outcome evaluation as part of a large-scale evaluation of this safety measure. The main objective of the outcome evaluation was to examine the impact of Nova Scotia's interlock program on participants and to help identify areas for improvement. More precisely, the goals of the outcome evaluation were:

- To determine the effectiveness of the program to reduce drink driving when combined with counselling and other Addiction Services components provided to the offender;
- To identify potential improvements to the program or implementation of the program;
- To determine the use of the program, e.g., participation rates and attrition.

Methodology

The outcome evaluation addressed the following questions:

1. How many participants re-offend, and how often, while enrolled in the program?
    a. How many were caught and convicted of drink driving while in the program? How many were arrested but not convicted; how many were caught for other driving-related offences?
    b. How many self-reported that they drove while drinking (or within an hour of drinking) while in the program?
2. How many failed attempts were logged on the interlock device?
    a. What were the reasons for the failed attempts?
    b. What was the BAC level of these failed attempts?
3. How many times did participants use the interlock device while in the program? What was the mileage driven during participation?
4. How many drove a non-interlock vehicle while in the program (based on self-reported data and conviction data)?
5. How many re-offend after they finished the program?
    a. How many are caught and convicted of drinking and driving?
    b. How many self-reported that they drove while drinking (or within an hour of drinking)?

6. What is the impact of the various aspects of the program, for example, voluntary versus mandatory participation?
7. Have participants' knowledge, attitudes and behaviors changed as a result of the program and in what ways?

The outcome evaluation also addressed interlock component-related research questions as discussed in the Process Evaluation Report (Robertson et al. 2010):

1. What is the distribution of participants in the program over time?
2. What is the attrition rate?
3. How do behavioral patterns among interlocked offenders change over time, more precisely with respect to blowing fails, violations and breath alcohol concentration (BAC) levels?
4. Is there a learning curve among participants on the device and does it change over time?
5. Is there a subpopulation that seems to be immune to the typical learning curve?
6. Is there a subpopulation that shows persistent and even deteriorating behavior over time?

Different types of data were used in this evaluation: conviction and crash records of individual participants, self-administered questionnaires to measure specific attitudes and behaviour, monthly counts of charges, convictions and crashes, and interlock logged events. For each type of experimental data (alcohol-related/interlock participants) except for the logged events, control data (no alcohol-related/no interlock participants) were also used to better support the findings. For the individual data analyses there were four different groups (see Table S-1): two experimental interlock groups (voluntary and mandatory interlock offenders) and two control non-interlock groups (offenders that had the option to participate in the interlock program and declined, and a retrospective control group consisting of offenders that would have been mandated into an interlock program had one existed).

Table S-1: Study groups

Experimental Groups	Control Groups
Voluntary interlock offenders: > Had an interlock device installed on their vehicle voluntarily.	Voluntary no-interlock offenders: > Had the option to participate in the interlock program, however declined.
Mandatory interlock offenders: > Were required to have an alcohol ignition interlock device installed on their vehicle.	Mandatory control group offenders: > Were charged with an alcohol-related offence and would have been required to install a device had an interlock program existed at the time.

The data were analyzed using five different methods:

1. Several descriptive sub-analyses were conducted to investigate demographic characteristics and other features to determine whether and how much participants in the control groups differed from participants in the experimental groups with respect to a variety of

dimensions. This provided important contextual information to ensure groups were sufficiently similar as well as to properly interpret results from any of the data analyses.

2. To provide insight into the effectiveness of the program in terms of recidivism, the longitudinal data from the experimental and control groups that were tracked over time were analyzed using a variety of survival analysis techniques. These techniques allowed the comparison of time to recidivate or crash across participants in each of the groups.
3. To bolster the findings from the survival analyses and to make the conclusions more robust, time series analyses techniques were used to study monthly counts of crashes, convictions and charges. This made it possible to study any potential effect of the implementation of Nova Scotia's interlock program on these counts.
4. Changes in attitudes and opinions regarding the interlock program, drinking behaviour, and drink driving behaviour, as measured by surveys, were investigated separately using regression analysis.
5. Finally, the interlock data were also analyzed without comparing these data to a control group to study behavioural trends of interlocked offenders.

The descriptive analyses revealed that in general there were no significant differences between the respective experimental and control groups at the beginning of the study period, meaning that these groups were well-matched and highly similar. In terms of alcohol-related charges, the control-voluntary group exhibited a recidivism rate of 8.9% during the study period, while the interlock-voluntary and interlock-mandatory had recidivism rates of 0.9% and 3% respectively after the installation of the interlock device. The recidivism rates for the interlock groups increased to 1.9% (voluntary group) and 3.7% (mandatory group) after the devices were removed from the vehicle, but they were still smaller than the rate for the control-voluntary group. This means that interlock participants were less likely to recidivate, even once the device was removed. In terms of alcohol-related crashes the control-voluntary group exhibited a recidivism rate of 1.6% during the study period, while the interlock-voluntary and interlock-mandatory groups had a 0.6% and 0.8% rate respectively. These differences in terms of crashes were not statistically significant.

Survival analysis. The results from the survival analyses demonstrated that:

> The interlock program was associated with a positive impact on reducing the risk for alcohol-related convictions of participants while driving.

> There seemed to be no difference between mandatory and voluntary participants in terms of risk for alcohol-related convictions.

> With respect to crashes, the analysis did not show a statistically significant difference between any of the studied groups.

Time series analysis. The results from the time series analyses suggested that:

> There were no permanent effects on the number of alcohol-related charges and convictions in the province as a whole associated with the implementation of the program.

- There were significant, albeit temporary effects in the first and seventh month after the program was implemented. These effects included:
  - a 13.32% decrease in the number of alcohol-related charges in September 2008; and,
  - a 9.93% decrease in the number of alcohol-related convictions in March 2009.
- With respect to crashes, time series analyses suggested that:
  - At the 5% level of statistical significance, there were no significant effects associated with the implementation of the program on the number of alcohol-related crashes with fatal and serious injuries.
  - There was a permanent effect at the 10% level of significance that represented a decrease of 0.0025 in the number of alcohol-related crashes every month since June 2009 (tenth month after the beginning of the program). Statistically speaking, this represented a small decrease (one fatal or serious alcohol-related crash in approximately 33 years). This was not unexpected as, to date most studies have not yet been able to definitively demonstrate a positive impact on crashes due to the small sample sizes and lack of sufficient data.

The small amount of data gathered from the questionnaires at exit and follow-up were insufficient to draw statistically significant conclusions and establish significant comparisons among groups with respect to changes in attitudes and opinions regarding the interlock program, drinking behaviour and drink driving behaviour. However, an interesting reported fact was that there was some evidence showing that a small proportion of interlock participants (in the mandatory group) drove a non-interlocked vehicle while in the program. This evidence should be considered in light of existing evidence about the alternative to interlocks, i.e., licence suspension, and which shows that many suspended drivers may drive anyway. While this finding may not be very surprising, it does speak to the importance of good monitoring of offenders while they are on the interlock, e.g., by tracking their mileage to help detect potential instances of driving non-interlocked vehicles.

The analysis of the interlock data suggested that there were learning curves which illustrated that offenders were more likely to violate at the beginning of program participation, but over time these violations decreased as offenders supposedly learned about, or experienced the consequences of program violations and the nuances associated with the functioning of, and compliance with, devices. In general, the curves were steepest at the beginning of program participation until approximately month 10, indicating that the learning effect may decrease or stop after a period of time. There were no large differences between male and female participants but there were clear differences between mandatory and voluntary participants. Although both groups revealed a learning effect, the effect was more pronounced for voluntary participants. In addition, clear differences were found between participants with condition 37 (condition on driver's license requiring a zero BAC) and participants without this condition. Although both groups revealed a learning effect, the effect was more pronounced for the participants without the condition.

Conclusion

In conclusion, with respect to specific deterrence (i.e., referring to preventing recidivism) there was strong evidence to suggest that participation in the interlock program reduced the risk of alcohol-related charges for the participants during the program. With respect to general deterrence (i.e., referring to a preventative effect on the entire population of drivers in Nova Scotia) there was a temporary decrease in the number of alcohol-related charges and convictions in the first and seventh month respectively following the implementation of the program. There was also some weaker evidence (i.e., at the 10% level of statistical significance) that there was a permanent decrease in the number of alcohol-related crashes with fatal and serious injuries every month since the tenth month after the beginning of the program.

When considering all the evidence combined, it can be argued that the implementation of the interlock program had a positive impact on road safety in Nova Scotia and that it reduced the level of drink driving recidivism in the province. There are also some promising indications to suggest a decrease in the number of alcohol-related road traffic crashes and fatalities due to the interlock program, although this finding should be confirmed with more data (crash data was available only until 2010). In sum, the evidence suggests the interlock program was better at preventing harm due to alcohol-impaired driving than the alternative of not using the interlock program.

Several recommendations were formulated based on the evidence from this study. These recommendations are:

- Continue the use of the interlock program in Nova Scotia;
- Consider the systematic use of a performance-based exit in the interlock program;
- Consider further strengthening of monitoring in the interlock program;
- Consider focusing on levels of risk in relation to non-compliance;
- Consider the continued monitoring of crash data.

TABLE OF CONTENTS

Executive Summary i

1. Background 1
   1.1 Alcohol ignition interlock programs 1
   1.2 Drinking and driving pattern in Nova Scotia 2
   1.3 Alcohol interlock program in Nova Scotia 2

2. Methodology 5
   2.1 Research questions 5
   2.2 Data 6
   2.3 Study design 10
   2.4 Data analysis 11

3. Descriptive analysis 15
   3.1 Demographic characteristics of the sample 15
   3.2 Individual data on convictions and crashes 18
   3.3 Conclusions 22

4. Survival analysis during and after the program 23
   4.1 Convictions 23
   4.2 Crashes 27
   4.3 Conclusions 30

5. Survival analysis after the program 33
   5.1 Convictions 33
   5.2 Crashes 37
   5.3 Conclusions 40

6. Time series analysis 43
   6.1 Charges 44
   6.2 Convictions 53
   6.3 Crashes 63
   6.4 Conclusions 71

7. Questionnaire data analysis 73
   7.1 Descriptive analysis 73
   7.2 Regression analysis 76
   7.3 Conclusions 77

8. Interlock data analysis 79
   8.1 Descriptive analysis 79
   8.2 Fails every three months period 81
   8.3 Fails in monthly periods 93
   8.4 Logistic regression analysis 101
   8.5 Conclusions 105

9. Overall conclusions 107

10. Recommendations 111

10.1 Continue the use of the interlock program in Nova Scotia 111

10.2 Consider the systematic use of a performance-based exit in the interlock program 111

10.3 Consider further strengthening of monitoring in the interlock program 112

10.4 Consider focusing on levels of risk in relation to non-compliance 112

10.5 Consider the continued monitoring of crash data 112

References 113

Appendix A: Offence Codes 115

Appendix B: Questionnaires 117

1. BACKGROUND

1.1 Alcohol ignition interlock programs

Alcohol interlocks have been commercially available for more than 30 years. The first alcohol interlock devices were developed in the 1960s as a tool to prevent drunk driving. These first devices were performance-based interlock systems, which required drivers to perform a perceptual or motor task designed to detect impairment prior to driving. While these devices were sensitive to individual variations in performance and impairment, they were incapable of discriminating between drivers with low to moderate breath alcohol concentration (BAC) levels. In the 1970s, new devices that were based on breath alcohol measurement were developed and proved to be considerably more reliable than the earlier performance-based devices. These devices were designed to incapacitate drunk driving offenders by preventing them from starting a vehicle when their BAC was in excess of a pre-set limit.

Across jurisdictions and around the world, the implementation of alcohol interlock programs to supervise impaired driving offenders is diverse. No two applications are alike – alcohol interlocks are applied with different purposes to different populations of users; users must meet different eligibility requirements; multiple agencies may be involved in administering these programs; and, their respective reporting, monitoring, and sanctioning features vary substantially. Of interest, the many different agencies involved in program delivery often have somewhat different roles and authority, and represent different systems (e.g., driver licensing system, enforcement system, adjudication system, health care system). As such, alcohol interlock programs are frequently based on collaborative initiatives that engage multiple agencies as partners in program delivery.

Despite the existence of alcohol interlock programs for more than two decades, jurisdictions continue to be challenged by the implementation of these programs. This has occurred because the development of effective policies, practices and procedures to support regulations has been ad hoc in many jurisdictions. To date, research has been unable to provide clear guidance on effective features of alcohol interlock programs, and, of greater concern, agencies have received limited guidance and support in relation to practices and procedures. Collectively, this has meant that the implementation of alcohol interlock programs has evolved using more of a trial and error process, and jurisdictions continue to modify and enhance existing protocols based on lessons learned.

Research is ongoing to identify the effective features of programs (for a comprehensive reference list see http://aic.tirf.ca/section2/references.php). At the same time, collaborative initiatives involving researchers, practitioners and government agencies are beginning to identify much-needed guidelines for programs based on existing knowledge and new experiences. In this regard, the process and outcome evaluation of Nova Scotia's interlock program can contribute to knowledge development.

1.2 Drinking and driving pattern in Nova Scotia

Although rates of impaired driving have decreased over the years, drinking and driving is still a significant problem in Nova Scotia (Alcohol Indicators Report 2011). From 2003 to 2007, 23.1% to 26.2% of drivers involved in crashes that caused serious injury had consumed alcohol. The Canadian rates for the same period ranged from 18.0% to 19.5%. In a 2007 telephone survey of driving practices and alcohol knowledge among young Nova Scotian men aged 19–35 years (Changing the Culture of Alcohol Use in Nova Scotia 2007), 46% reported driving within two hours of consuming alcohol at least once in the past 12 months. Of those, 40% reported doing this 1–2 times; 24% said 3–5 times; and 36% said 6 or more times.

According to the Alcohol Indicators Report 2011, heavy-drinking rates in Nova Scotia are high. In 2007–2008, 38.9% of males and 17.5% of females engaged in heavy monthly drinking. During the same time frame, 17.9% of males and 7.0% of females engaged in heavy weekly drinking. Heavy-drinking rates are particularly high among young adults; the usual consumption pattern for 51.7% of Nova Scotia university undergraduate students in 2004 was five or more drinks on the days they drank, with 27.2% of all university students drinking heavily at least once a week.

According to a student's survey (Student Drug Use Survey 2012), rates of drinking and driving are decreasing. In 2012, 4% of students in grade 7, 9, 10 and 12 reported having driven a vehicle within an hour of consuming two or more alcoholic drinks. The drinking and driving rate among all students was 5.3% in 2007 and 6.6% in 2002. Among students in grades 10 and 12 with a driver's license, 10% drove within an hour of consuming two or more drinks in 2012 compared to 13.6% who reported doing so in 2007 and 14.8% in 2002. No gender or location differences were observed. In 2012, 5.4% of students in grades 7, 9, 10 and 12 were in a motor vehicle accident with them as a driver and less than 1% of students reporting drinking and driving prior to their accident. In 2012, 16.8% of students were a passenger in a vehicle with a driver who was impaired, compared to 19.2% in 2007, and 22.8% in 2002. No differences were observed for gender, grade level or school location.

Overall, per capita alcohol consumption among Nova Scotians increased 6.6% during a 20-year period, growing from 7.6 litres of pure alcohol in 1991 to 8.1 litres in 2010. The alcohol-related mortality rate increased by 27% between 2002 and 2008 (Alcohol Indicators Report 2011).

1.3 Alcohol interlock program in Nova Scotia

In September 2008 Nova Scotia's Alcohol Ignition Interlock Program was implemented. The overall objective of the Alcohol Ignition Interlock Program (AIIP) was to improve road safety and reduce the number of road traffic crashes and fatalities that may occur due to impaired driving.

Nova Scotia's program involves both voluntary and mandatory components. It is voluntary for first-time offenders deemed to be a 'low' or 'medium' risk (as determined by Addiction Services of Nova Scotia through the Alcohol Rehabilitation Program) and mandatory for those who are deemed to be a 'high' risk and/or those convicted of drinking and driving (or the refusal of the breathalyzer) more than once in the past ten years. After entering the program, participants must have an alcohol interlock device installed on their vehicle(s). They will then receive an interlock licence, and must participate in ongoing rehabilitation counseling sessions throughout the interlock period. The licence allows them to drive an interlock-equipped vehicle during their revocation period as long as they are compliant with the terms of the program. Their licence is stamped with an 'R' indicating that they are restricted to driving an interlock-equipped vehicle. Furthermore, participants are only permitted to operate specific interlock-equipped vehicles (e.g., a participant is not allowed to operate another participant's interlock-equipped vehicle) and the Registry of Motor Vehicles (RMV) must be notified of all vehicles in which an interlock device is installed.

In April 2010 TIRF finalized a process evaluation as part of a large-scale evaluation of this safety measure (Robertson et al. 2010). The overall objective of the process evaluation was to obtain a common understanding about how Nova Scotia's interlock program was developed and how it was implemented in order to identify potential areas for improvement. Overall, the results revealed that the implementation of the alcohol interlock program in Nova Scotia proceeded according to the plan. "While some adjustments were required during program implementation to adapt to a changing environment, some instances of incompatible processes, and to address miscommunication, overall the implementation was highly consistent with the plan that was developed to guide this initiative" (Robertson et al. 2010, page 51). The second phase of the Alcohol Ignition Interlock Program evaluation is the outcome evaluation described in this report.

2. METHODOLOGY

This outcome evaluation was conducted by the TIRF as part of a large-scale evaluation of the alcohol ignition interlock program in Nova Scotia. The overall objective of the outcome evaluation was to examine the impact of the program on participants and to help identify areas for improvement. More precisely, the goals of the outcome evaluation were:

- To determine the effectiveness of the program in reducing drink driving when combined with counselling and other Addiction Services components provided to the offender;
- To identify potential improvements to the program or implementation of the program; and,
- To determine the use of the program, e.g., participation rates and attrition.

To evaluate the impact of Nova Scotia's interlock program on participants and help identify areas of improvement, different types of analyses were used. In particular survival analysis and interrupted time series analysis were used to evaluate the impact of the program. In this section the methodology of the different analyses is described.

2.1 Research questions

The outcome evaluation addressed the following questions:

6. How many participants re-offend, and how often, while enrolled in the program?
   a. How many were caught and convicted of drink driving while in the program? How many were arrested but not convicted; how many were caught for other offences?
   b. How many self-reported that they drove while drinking (or within an hour of drinking) while in the program?
7. How many failed attempts were logged on the interlock device?
   a. What were the reasons for the failed attempts?
   b. What was the BAC level of these failed attempts?
8. How many times did participants use the interlock device while in the program? What was the mileage driven during participation?
2. How many drove a non-interlock vehicle while in the program (based on self-reported data and conviction data)?
3. How many re-offend after they finished the program?
   a. How many are caught and convicted of drinking and driving?
   b. How many self-reported that they drove while drinking (or within an hour of drinking)?
9. What is the impact of the various aspects of the program, for example, voluntary versus mandatory participants?

10. Have participants' knowledge, attitudes and behaviours changed as a result of the program and in what ways?

The following additional interlock related questions were answered in this outcome evaluation (these are interlock component related research questions 6 through 11 in the process evaluation framework):

11. What is the distribution of participants in the program over time?
12. What is the attrition rate?
13. How do behavioural patterns among interlocked offenders change over time, more precisely with respect to blowing fails, violations and breath alcohol concentration (BAC) levels?
14. Is there a learning curve among participants and does it change over time?
15. Is there a subpopulation (by gender, by mandatory/voluntary condition) that seems to be immune to the typical learning curve?
16. Is there a subpopulation that shows persistent and even deteriorating behaviour over time?

2.2 Data

Information from different existing data sources was used in this evaluation. This included driver licensing information, crash data and conviction data. These data were obtained from a variety of sources:

- Registry of Motor Vehicles' data: Driver licensing information, individual crash and conviction;
- TIRF Fatal and Serious Injured Crash data bases (Monthly crashes);
- Justice Nova Scotia (monthly charges and convictions);
- Interlock data (Alcohol Countermeasure System Corp.); and,
- Questionnaire data (Addiction Services Nova Scotia, Opinion Search Inc.).

2.2.1 Individual driver data

In order to evaluate the effectiveness of Nova Scotia's Alcohol Ignition Interlock Program, data from two experimental groups (voluntary and mandatory interlock offenders) and two control groups (non-interlock offenders recruited during the current intake period and non-interlock offenders recruited retrospectively) were compared across several measures. For this study the classification of interlock participants as mandatory or voluntary was based on a proxy measure using the anticipated termination date of the program. If the time between the interlock device installation date and the anticipated termination date was less than a year, then the participant was considered voluntary, otherwise the participant was considered mandatory.

At the most basic level, comparisons were made between those who participated in the interlock program and those who did not with respect to driving history, crash records, and conviction records (see appendix A for a description of the offence codes considered). This was to determine whether drivers who participated in the interlock program went on to have fewer alcohol-related crashes, convictions, and problems. A more detailed illustration of each group and the comparisons is presented below:

Figure 2-1: Study groups definition:

Experimental Group 1: Voluntary interlock

- Had an interlock device installed on their vehicle voluntarily from March 2010 to December 2012.
- Driving, conviction, and crash records to be tracked from **eighteen months prior** to the voluntary installation of the interlock device, until June 30th, 2014.

Vs.

Control Group 1: Voluntary no-interlock

- Had the option to participate in the interlock program, and declined. Registration from March 2010 to December 2012.
- Driving, conviction, and crash records to be tracked from **eighteen months prior** to registering to be in the control group, to June 30th, 2014.

Experimental Group 2: Mandatory interlock

- Were required to have an alcohol ignition interlock device installed on their vehicle from March 2010 to December 2012 as a result of an alcohol-related driving offence.
- Driving, conviction, and crash records to be tracked from **eighteen months prior** to the mandatory installation of the interlock device, to June 30th, 2014.

Vs.

Control Group 2: Mandatory control group

- Were charged with an alcohol-related offence, between March 2003 and December 2005, that *would have* mandated them into an interlock program, had one existed at the time (see below for a list of these offences). That is, had they been convicted of the *same offence* in 2008, they would have been required to use an ignition interlock device.
- Driving, conviction, and crash records will be analysed from **eighteen months prior** to the alcohol-related driving offence that would have mandated participating in the AIIP, to June 30th, 2007.

Other detailed comparisons were made with respect to specific attitudes of participants including attitudes towards alcohol, drinking and driving and interlock logged events.

Questionnaire information among participants in the study. For each of the experimental groups, information was collected at three different points in time; during intake, exit and at a six month follow-up after exiting the program. For the control group (that was composed during the current intake window) information was collected at two different points in time; intake and exit to the Driving While Impaired (DWI) program. The information was gathered using self-administered questionnaires on paper (see appendix B) and included:

> demographics;
> self-reported behaviour;
> readiness to change;
> Research Institute on Addictions Self Inventory (RIASI); and,
> expectations about Interlocks.

The table below summarizes what information was collected at each point:

Table 1: Overview of data collection tools and data collection scheme

	Intake interview	Exit interview	6-month follow-up
Demographic information	Exp. — volunteers Exp. — mandatory Control — current	Exp. — volunteers Exp. — mandatory o Control — current	Exp. — volunteers Exp. — mandatory
Readiness to change	Exp. — volunteers Exp. — mandatory Control — current	Exp. — volunteers Exp. — mandatory Control — current	Exp. — volunteers Exp. — mandatory
Attitudes about interlocks	Exp. — volunteers Exp. — mandatory Control — current	Exp. — volunteers Exp. — mandatory Control — current	Exp. — volunteers Exp. — mandatory
Research Institute on Addictions Self Inventory	Exp. — volunteers Exp. — mandatory Control — current	Exp. — volunteers Exp. — mandatory Control — current	Exp. — volunteers Exp. — mandatory
Self-reported behaviour	Exp. — volunteers Exp. — mandatory Control — current	Exp. — volunteers Exp. — mandatory Control — current	Exp. — volunteers Exp. — mandatory

Demographics. This questionnaire contained standard demographic questions. The information gathered included gender, age, marital status, employment status, available cars, recidivism status and interlock knowledge.

Self-reported behaviour. This questionnaire was about drinking and driving behaviour. At intake the questionnaire had three questions about recent drinking and driving situations and the exit and follow-up questionnaires included other questions related to drinking and driving during participation in the program and about future expected behaviour.

Readiness to change. This questionnaire contained four subscales, each of which corresponded to one of Prochastka and DiClemente's (1986) stages of change: pre-contemplation, contemplation, action and maintenance based on their model of behaviour change. The subscale with the highest score represented the participant's current stage in this model of behaviour change. Pre-contemplation refers to a stage in which the individual is not considering a change in their behaviour. In the contemplation stage the individual thinks about changing their behaviour. In the action stage the individual is actively changing their behaviour and, finally, individuals who have reached the maintenance stage are working to prevent a relapse.

Research Institute on Addictions Self Inventory (RIASI). This is an empirically-derived instrument specifically designed for use with drink drivers (Nochajski et al. 1994). In addition to providing an indication of the extent of alcohol use, it has a subscale that assesses the likelihood of a repeat offence.

Expectations about interlocks. This questionnaire measures what clients expect from their participation in the ignition interlock program.

Interlock Information. Interlock data were provided by Alcohol Countermeasure System Corp. (the sole vendor in Nova Scotia). The data contained information for each participant in the interlock program, the date of installation of the interlock device, a list of events with the date/time of each interlock event and the type of the event during their participation in the program. The events were the results of the breath sample tests when trying to start the car (at start-up) or after having started the car (running retest). Results from these breath samples were classified according to the BAC level as "pass" (BAC level under 0.02%), "fail" (BAC level above 0.02%). The exact BAC level at each event was also provided.

2.2.2 Monthly counts data

Monthly counts of alcohol-related charges, convictions and crashes in Nova Scotia were compared with non-alcohol-related charges, convictions and crashes in Nova Scotia from 1998 to 2013 (see appendix A for a description of the offence codes considered).

The evaluation period included information from approximately ten years (from 1998 to 2008) before Nova Scotia implemented the interlock program and approximately five years (from 2008 to 2013) after implementation. Note however that crash data were only available up to 2010 (two years after AIIP implementation). Three different experimental time series were produced:

- alcohol-related charges: monthly counts, 1998-2013;
- alcohol-related convictions: monthly counts, 1998-2013; and,
- alcohol-related fatal and serious crashes: monthly counts, 1998-2010.

For each of the alcohol-related time series, a corresponding non-alcohol-related control time series was included in the analyses to model the possible impact of AIIP after its implementation as well as to control for possible confounding variables that may have affected the experimental and control counts. The non-alcohol-related control time series were designed to be similar to the experimental series except for their relationship with alcohol, and as such, they should not have been influenced by the implementation of the alcohol ignition interlock program. (The purpose of the control time series was to eliminate alternate explanations of the possible results.) This made it possible to determine whether the implementation of the program was associated with any differences between the experimental and control data.

Other counts during the study period were also obtained and included in the analyses to control for their possible impact.

- Population: Population estimates by quarters, aged 16 and over, 1998 to 2013 (Statistics Canada, 2014);
- Unemployment rate: Monthly percentage of adults aged 15 and over in the labour force that are unemployed (Statistics Canada, 2014);
- Heavy drinking: Annual population aged 12 and over who reported having 5 or more drinks on one occasion, at least once a month in the past 12 months, 1998 to 2012 (before 2008 available information was biannual) (Statistics Canada, 2013); and,
- Alcohol sales: Average litres bought annually by adults aged 15 and older, 1998-2013 (Statistics Canada, 2014), Statistics Canada, CANSIM Table 183–0019 (per capita consumption estimates determined using population aged 15 years and over).

2.3 Study design

The study design utilized in this evaluation was a longitudinal study whereby data from two experimental groups (voluntary and mandatory interlock offenders) and two control groups (non-interlock offenders recruited during the current intake period and non-interlock offenders recruited retrospectively) were compared across several measures. The longitudinal data from the experimental and control groups that were tracked over time were analyzed using a variety of survival analysis techniques. This allowed for the comparison of the behaviour of interlocked offenders (voluntary or mandatory) with non-interlocked offenders in the control groups, in order to draw conclusions about the true impact of the program. Accounting for the potential influence of other possible factors was achieved by including as many control variables as possible based on the information gathered during pre-determined times (e.g., demographic information, information about readiness to change, etc.). Survival analysis made it possible to distinguish between short-term and long-term safety effects by using time until an event occurred as useful information in the analyses. Note that survival analysis is a very flexible technique that accounts for situations when the timing of the delivery of an intervention differs across individuals. This type of analysis provided answers to research questions 1, 4, 5 and 6.

Time series analyses were also used to bolster the findings from the survival analyses. It made it possible to obtain a better understanding of long-term trends by controlling for factors that may have influenced the results such as population, unemployment, heavy drinking and alcohol sales. These analyses were performed by looking at monthly rates of alcohol-related crashes, convictions and charges over a longer tracking period, including a period before and after the implementation of Nova Scotia's interlock program. These monthly counts were compared to similar counts of non-alcohol-related events. The evaluation period was approximately ten years (from 1998 to 2008) before Nova Scotia implemented the interlock program and approximately five years (from 2008 to 2013) after implementation.

The differences between the survival analyses and time series analyses may be explained by specific and general deterrent effects of the interlock program. In particular, the survival analyses made it possible to examine the impact of the program on specific, individual participants by looking at the behaviour of drink driving offenders enrolled in the interlock program in comparison with those who did not participate in the program. On the other hand, the time series analyses used data from the province as a whole, including drivers who were not enrolled in the interlock program. As such, the time series analyses enabled permitted the study of the general impact of the interlock program on the general population.

Changes in knowledge, attitudes and opinions regarding the interlock program, drinking behaviour, and drink driving behaviour were also investigated separately by comparing scores coming from the questionnaires that were administered at predetermined times. Theses analyses were used to answer research questions 1b, 4, 5b, and 6. Regression analysis was used to determine if the changes in the different behavioural scores were significant among the different groups taking into account possible factors such as demographics and initial behavioural scores. Useful findings from these analyses were integrated with findings from previous analyses to further elaborate on the conclusions. This generated information was useful to help answer research question 7.

Finally, the interlock data were analyzed without comparing them to behavioural indices from a control group. It warrants mentioning that the interlock device logs a huge amount of data (approximately 1.5 megabytes of information per offender per year on the interlock). Indexing techniques were used to gauge how many failed attempts were logged on the interlock device, the reasons for the failed attempts, and the number of tests delivered per offender. This analysis was used to answer research questions 2 and 3 and the additional interlock component-related research questions.

2.4 Data analysis

Data analysis was conducted in five different ways using Stata/MP 13.1 for Windows 64-bit x86-64 (StataCorp., 2013; Cleves et al, 2004; Becketti, 2013). First, comparisons about demographic, convictions and crash data are performed using descriptive and bivariate analyses. Then, survival analyses were conducted to evaluate the true impact of the program on convictions and crashes. The survival function after inclusion date (interlock device installation date for interlock groups, consent date for the voluntary control and offense date for the mandatory control groups) is estimated using Kaplan-Meier estimators. The Kaplan-Meier estimator is nonparametric, which requires no parametric assumptions and provides estimates of the probability of surviving to (or being free of the event in question) at different times. A graph of the survival function provides a summary of the time-related information and it is possible to compare data from different groups by visual inspection of their respective estimated survival time. Cox proportional hazard regression was used to consider additional covariates in the survival models. Furthermore, the model helps to determine how significant the hazard ratios are. Cox proportional hazard models assume that the hazard ratio is constant over time; consequently tests of proportional-hazards assumption were used in the analyses. In case that the proportional hazard assumption does not hold, flexible parametric models were used as an alternative to the semi-parametric Cox model (Royston and Parmar, 2001).

Interrupted S-ARIMA(X) time series analysis was also used to evaluate the impact of the program (Linden and Adams 2010). Using data only from the experimental group in the period preceding the implementation of the AIIP (i.e., January 1998 through to September 2008, inclusive), the structure of the experimental time series was investigated and used to build the final ARIMA time series model, as suggested by McCleary and Hay (1980). This approach is appropriate when there are a sufficient number of data points in the pre-intervention period (approximately 60 data points), corresponding to five years' worth of data. This approach is recommended because using the entire series, including the post-intervention data can obscure the true structure of the time series due to the potential impact of the intervention.

These pre-intervention series of counts were investigated with special attention given to the overall pattern, outliers and variance of the data. When there were outliers or when there was non-stationary variance, pre-intervention time series were transformed using the natural log transformation to mitigate their impact. Note that this transformation was not always used because the numbers in different time series were too low. Trends were investigated to see if local differencing of the pre-intervention series was required (if the trend was significant the time series was locally differenced to make the series stationary). Seasonality was investigated using a correlogram to see if seasonal differencing of the pre-intervention series was required (if seasonality was significant the time series were seasonally differenced). The impact of outliers was further reduced by including a dummy variable for each outlier when its Z-score was greater than 2.5.

Selection of the final pre-intervention model was based on a comparison of the information criteria values of potential models, along with ARMA terms that were significant as well as within the bounds of stationarity and invertibility (see Yaffee, 2000). Tests of white noise were used to ensure the final model's residuals were white noise and tests of normality were used to ensure the residuals were normally distributed. Also, robust standard errors were used when modeling the data. Once the final ARIMA model was found based on the experimental pre-intervention time series using the approach described above, a set of dummy variables to model the intervention along with control group data and series of population estimates and alcohol consumption for the entire study period were inserted in the final model simultaneously to test the hypotheses about possible intervention effects of AIIP. Effects of these dummy variables are described using adjusted monthly percentage changes for log-transformed series (coefficients of the log transformed data in the final model are transformed using the number 'e'). Effects were considered significant if p-values were smaller than 0.05 (i.e., 5% level).

Using this approach, three different model structures were tested. In the first model, the sudden permanent model, the dummy variables representing the implementation of AIIP are used. This model assumes the impact of the implementation of AIIP is immediate and permanent. In the second model, the gradual permanent model, an interaction effect between these dummy variables and the post-implementation/post-change trend is used. This model assumes there is a permanent change, but the change is gradual and not sudden (see Linden and Adams 2010). The third model, the sudden temporary model assumes that there was an impact, but the impact did not last and can be represented by a spike in the data associated with the implementation or change. This model was tested using a dummy variable that represents a pulse (the pulse variable is given the value 1 for the month representing the implementation or change and the value 0 for all other months). Because it is not known a priori in which month the impact in any of the three models would occur, thirteen different months have been tested: the month of the implementation or change itself; and twelve months following the implementation or change. Note that this approach accounts for the possibility that the program impact is only effective after some time, for example because there might be a delay between the passing of the legislation and the enrolling of the first program participants, or because it may take some time before enough participants have enrolled for the program to potentially have any impact.

Data from questionnaires were used to compare the groups at the beginning of the study and their change over time. Descriptive and bivariate statistics were used to describe the groups in terms of alcohol consumption and drinking-driving behaviours. Results were presented at different moments of time, at intake interview, at exit and follow-up. Regression analyses were used to study the change in the RIASI total and recidivism scores and determined if there are significant differences between the groups. Possible confounding such as demographics and initial behavioural scores were also included in the regression models.

Interlock data analysis was conducted to determine the use of the program and understand behavioural patterns of offenders on an interlock. The analysis examined events logged since the implementation of the program in 2008 until July 2014, for participants that have the interlock device installed until December 2012 (the intake period). Behavioural patterns were investigated in time blocks of three months (in tables) and one month (in figures) to reveal changes over time. Although the maximum time on the interlock in our sample was 68 months, approximately 75% of the participants were in the program for only 30 months. As such, the tracking period used was 30 months, individualized per participant since their device installation date. Descriptive statistics including counts and percentages, along with 95% confidence intervals (95%-CI) have been calculated. The data have been analyzed in relation to several different types of events. These events include: blowing a breath sample over 0.02, blowing a breath sample over the provincial limit of 0.05, blowing a breath sample over the criminal limit of 0.08, start-up violations and running retest violations. The analyses have also been broken down by gender as well as mandatory versus voluntary participants. Logistic regression analysis allowed comparing the odds of a failed test in the interlock device to the odds of a passed test while simultaneously controlling for several factors like months in the program, gender, mandatory/voluntary, start-up/running type of test, condition 37 (condition on driver's license requiring a zero BAC) and average mileage driven.

Finally, in order to protect identity and in compliance with the Privacy Impact Assessment of this study, throughout the report only percentages are reported in descriptive tables where the cell counts were low.

3. DESCRIPTIVE ANALYSIS

3.1 Demographic characteristics of the sample

The sample contained 929 interlock experimental and 359 non-interlock control participants. The study groups (interlock-mandatory and interlock-voluntary versus control-mandatory and control-voluntary) correspond to the definition given in the methodology section (see Figure 2-1). The data from the interlock groups were obtained from Alcohol Countermeasure System Corp. and the data for the control groups were obtained from Addiction Services Nova Scotia and the Registry of Motor Vehicles.

Table 3-1 below provides a brief description (in percentages) of the demographic characteristics for the study groups in terms of gender, age and condition 37. Condition 37 (condition on driver's license requiring a zero BAC) is a discretionary decision of the Registrar of motor vehicles. It could be based on the risk rating of the offender or the opinion of medical practitioner and the usual term is for three years.

Table 3-1: Demographic Characteristics of the study groups, Interlock (mandatory and voluntary) and Control (mandatory and voluntary), in percentages

	Interlock-M	Interlock-V	Control-M	Control-V	Tests
Gender					$p(F)=0.001$ $p_1(\chi^2)=0.025$ $p_2(F)=0.54$ $p_3(\chi^2)=0.11$ $p_4(F)=0.48$
Female	9.4	14.5	12.1	18.9	
Male	90.6	85.5	87.9	81.1	
Age					$p(F)=0.000$ $p_1(\chi^2)=0.002$ $p_2(F)=0.000$ $p_3(\chi^2)=0.000$ $p_4(F)=0.05$
15-24	4	7.1	30.3	22.4	
25-34	27.2	17.8	18.2	26.5	
35-44	26.6	23.3	24.2	18.7	
45-64	37.8	45.7	15.2	29	
65+	4.4	6.1	12.1	3.4	
Condition 37					$p(F)=0.000$ $p_1(\chi^2)=0.000$ $p_2(F)=0.000$ $p_3(\chi^2)=0.023$ $p_4(F)=0.04$
Have	69.8	22.4	3	17.9	
Do not have	30.2	77.6	97	82.1	

A series of tests have been conducted to analyze the observed differences between the groups. The Pearson's chi-squared test (X^2) was used to compare the observed frequencies for the groups when the frequencies were sufficiently large (above 5); otherwise the Fisher's exact test was used instead. The p-values of these tests are presented in the table. The symbols $p(\chi^2)$ and p(F) represent the p-values for the chi-squared test and the Fisher's exact test respectively. When no subscript is used the comparison was between the four groups. A subscript represents the comparison between two specifics groups in the following way:

1. Interlock-mandatory/ interlock-voluntary
2. Interlock-mandatory/ control-mandatory
3. Interlock-voluntary/ control-voluntary
4. control-mandatory/ control-voluntary

With respect to gender, overall the distribution of gender is dependent on the group, but the only paired comparison that revealed dependency (p-value=0.025) was the one between both interlock groups, where the interlock-voluntary group has a larger percentage of females than the interlock-mandatory group. In all other paired comparisons, the null hypothesis of independence between gender and group could not be rejected (p-values>0.1). Overall, in all groups there is higher a percentage of males than females. However, even in the interlock groups where the gender and group are dependent, the differences in proportions of males and females are not too large.

With respect to age, overall and for all paired comparisons the distribution of age categories is dependent on the group. The control groups have the higher percentages (30.3% and 22.3%) of participants in the youngest age group 15-24. The interlock groups have only 4% and 7.1% of participants in the same young age group.

With respect to condition 37 the observed frequencies are significantly different overall and for all paired comparisons. The mandatory interlock group has the highest percentage (69.85%) of drivers with this condition.

Other comparisons were possible based on the demographic questionnaires (see table below) for all contemporary groups (this excludes the control-mandatory due to the fact that this is the retrospective control group for which these data were not available). However, for these comparisons the sample size was smaller due to low response rates in the questionnaires. The sample contains 163 interlock and 318 non-interlock participants. The only significant results were with respect to the number of available vehicles (the control-voluntary has a larger percentage of participants with no vehicles than the interlock-voluntary) and with respect to being first-time offenders (not surprisingly the interlock-voluntary group has a larger percentage of first-time offenders than the interlock mandatory).

Table 3 2: Demographic Characteristics of the study groups based on questionnaires, Interlock (mandatory and voluntary) and Control voluntary, in percentages

	Interlock-M	Interlock-V	Control-V	Tests
Gender				$p(\chi^2)=0.468$
Female	13.5	18.46	19.2	$p_1(\chi^2)=0.4$
Male	86.5	81.54	80.8	$p_3(\chi^2)=0.897$
Age				
15-24	2.2	10.8	22.5	
25-34	30.3	26.1	26.7	$p(\chi^2)=0.001$
35-44	28.1	20	19.1	$p_1(F)=0.133$
45-64	36	35.4	28.2	$p_3(\chi^2)=0.144$
65+	3.4	7.7	3.5	

Table 3 2: Demographic Characteristics of the study groups based on questionnaires, Interlock (mandatory and voluntary) and Control voluntary, in percentages				
	Interlock-M	**Interlock-V**	**Control-V**	**Tests**
Marital-status				
Single	37.36	38.03	47.9	$p(\chi^2)=0.104$ $p_1(F)=0.79$ $p_3(F)=0.12$
Married	25.27	19.72	16	
Living together	13.19	11.27	16.3	
Divorce/separate	18.7	26.8	14	
Widow	3.3	1.4	1.3	
Other	2	2.8	4.6	
Employment status				
Employed	74.2	80.6	67	$p(F)=0.541$ $p_1(F)=0.86$ $p_3(F)=0.224$
Unemployed	12.4	9	15.8	
Retired	5.6	4.5	6.5	
Other	7.9	6	10.6	
Other with lic. in household				
Partner	38.9	42.9	32.6	$p(\chi^2)=0.065$
Children	13.9	10.2	4.9	
Sibling	0	4.1	5.2	
Friend	5.6	2	6.7	
Other relative	15.3	22.4	28.1	
Other	4.2	0	2.6	
None	22.2	18.4	19.8	
Other restricted by interlock in household				
Yes	2.7	3.8	2.3	$p(F)=0.7$ $p_1(F)=1$ $p_3(F)=0.62$
No	97.3	96.2	97.7	
Vehicles available				
None	19.1	5.9	23	$p(F)=0.032$ $p_1(F)=0.064$ $p_3(F)=0.05$
1	50.6	50	43.9	
2	20.2	29.4	23.6	
3 and over	10.1	14.7	9.5	
First-time offender				
Yes	28.9	88.6	81	$p(\chi^2)=0.000$ $p_1(\chi^2)=0.000$ $p_3(\chi^2)=0.137$
No	71.1	11.4	19	
ethnic				
White	92.3	94.2	93.4	$p(F)=0.79$ $p_1(F)=0.94$ $p_3(F)=0.95$
Black	2.2	2.9	3.6	
Indian	3.3	1.4	1	
Other	2.2	1.4	2	

In sum, the descriptive data show no relevant differences between the experimental (interlock) and their respective control group at the beginning of the study with the exception of age. This information is pertinent to the interpretation of any findings in the multivariate analyses, notably the survival analysis.

3.2 Individual data on convictions and crashes

Relevant information in terms of convictions and crashes for each participant in the study groups were obtained from the Registry of Motor Vehicles. Conviction data are all alcohol related (see appendix A for the offense codes). From the data it was possible to identify 91.1% of all participants in the study.

The figure below shows the percentage of participants in each group that have convictions in the 18 months before their inclusion date (interlock device installation date for the interlock groups, registration date in the group for the voluntary-control and offense date for the mandatory-control group). The differences in percentages of convictions may be biased due to the differences in the inclusion dates for the following reason. All participants are alcohol-related offenders, but the time since their alcohol-related offense and the inclusion date may be larger than 18 months in which case it is not counted in this study.

Figure 3-1: Percentages of participants with and without alcohol-related convictions before their inclusion date per group

convict_before	group: mand-inte	vol-inter	vol-contr	mand-cont	Total
do not have	63.82	8.13	61.16	0.00	46.34
have	36.18	91.88	38.84	100.00	53.66
Total	100.00	100.00	100.00	100.00	100.00

Pearson chi2(3) = 309.5751 Pr = 0.000

Figure 3-2 shows the percentage of participants in each group that have convictions after their inclusion date during the tracking period (up to June 2014 for the contemporary groups and June 2007 for the retrospective group).

Figure 3-2: Percentages of participants with and without alcohol-related convictions after their inclusion date per group

convict_after	group: mand-inter	vol-inter	vol-contr	mand-contr	Total
do not have	96.98	99.06	91.07	96.97	96.42
have	3.02	0.94	8.93	3.03	3.58
Total	100.00	100.00	100.00	100.00	100.00

Pearson chi2(3) = 25.6354 Pr = 0.000

The group with smaller percentage (0.9%) of participants with alcohol convictions is the voluntary-interlock and the group with a larger percentage (8.9%) of participants with alcohol-related convictions is the voluntary-control. Since the data for the mandatory-control group are very limited this group is not included in further analyses in the next sections.

The figure below shows similar information but for the interlock groups only those who had an interlock device removed during the intake period (March 2010-December 2012) are considered. The table shows for the interlock groups, percentages of participants having alcohol-related convictions after the interlock had been removed. Note that the percentages are larger after removing the device, however they are still smaller than the percentages of participants having convictions in the control-voluntary group.

Figure 3-3: Percentages of participants with and without alcohol-related convictions after their inclusion date for control group and after removal of interlock device for interlock groups

convict_after	group: mand-inter	vol-inter	vol-contr	Total
do not have	96.27	98.11	91.07	95.57
have	3.73	1.89	8.93	4.43
Total	100.00	100.00	100.00	100.00

Pearson chi2(3) = 16.6386 Pr = 0.000

Crash data were collected for the contemporary groups (all except the mandatory-control group) however, some data were not alcohol related. Figure 3-4 shows the distribution of type of crashes with respect to alcohol by each group, before and after the participant's inclusion date.

Figure 3-4: Percentages of types of crashes for participants per group before and after their inclusion date

Before inclusion date

crash_alcohol	mand-inte	vol-inter	vol-contr	Total
impaired by alcohol	48.28	34.38	68.42	47.50
Suspected use of alcohol	17.24	28.13	5.26	18.75
non-alcohol related	34.48	37.50	26.32	33.75
Total	100.00	100.00	100.00	100.00

Pearson chi2(4) = 6.7424 Pr = 0.150
Fisher's exact = 0.162

After inclusion date

crash_alcohol	group: mand-inte	vol-inter	vol-contr	Total
impaired by alcohol	8.33	6.45	11.76	8.76
Suspected use of alcohol	0.00	3.23	5.88	2.19
non-alcohol related	91.67	90.32	82.35	89.05
Total	100.00	100.00	100.00	100.00

Pearson chi2(4) = 4.6319 Pr = 0.327
Fisher's exact = 0.256

The number of alcohol-related crashes (impaired and suspected) decreased after the inclusion date. Again, the group with a smaller percentage of alcohol-impaired crashes during the tracking period is the voluntary-interlock (6.5%) and the group with a larger percentage (11.8%) of alcohol-impaired crashes is the voluntary-control group. Note however, that the independence tests have p-values greater than 0.05, thus the hypothesis of independence of types of crashes and the different groups cannot be rejected. Therefore, these results are not statistically significant.

The figure below shows the percentage of participants in each group that have crashes before their inclusion date. The first part of the figure shows all type of crashes and the second part shows only alcohol-related crashes. Before the inclusion date, 4.1% of participants have alcohol-related crashes.

Figure 3-5: Percentages of participants with and without crashes before their inclusion date per group. Top: all crashes, bottom: alcohol-related crashes

crashes_before	group mand-inte	vol-inter	vol-contr	Total
do not have	95.38	90.09	94.79	93.86
have	4.62	9.91	5.21	6.14
Total	100.00	100.00	100.00	100.00

Pearson chi2(2) = 10.8735 Pr = 0.004

alc_crashes_before	group mand-inte	vol-inter	vol-contr	Total
do not have	97.19	93.81	95.71	95.94
have	2.81	6.19	4.29	4.06
Total	100.00	100.00	100.00	100.00

Pearson chi2(2) = 6.2588 Pr = 0.044

The figure below shows the percentages of participants in each group that have crashes after their inclusion date. The first part of the figure shows all type of crashes and the second part shows only alcohol-related crashes. After the inclusion date, 0.88% of participants have alcohol-related crashes. The group with a smaller percentage (0.62%) of participants with alcohol-related crashes during the tracking period is the voluntary-interlock and the group with larger percentage (1.6%) of participants with alcohol-related crashes is the voluntary-control. However, the p-values of the independence test are greater than 0.05, therefore, these results are not statistically significant.

Figure 3-6: Percentages of participants with and without crashes after their inclusion date per group. Top: all crashes, bottom: alcohol-related crashes

crashes_after	group: mand-inte	vol-inter	vol-contr	Total
do not have	89.93	91.95	90.57	90.68
have	10.07	8.05	9.43	9.32
Total	100.00	100.00	100.00	100.00

Pearson chi2(2) = 1.0210 Pr = 0.600

alc_crashes_after	group: mand-inte	vol-inter	vol-contr	Total
do not have	99.17	99.38	98.43	99.12
have	0.83	0.62	1.57	0.88
Total	100.00	100.00	100.00	100.00

Pearson chi2(2) = 0.7255 Pr = 0.696
Fisher's exact = 0.791

The figure below shows percentages of interlock participants having crashes after the interlock had been removed. The percentages of alcohol related crashes are larger after removing the device, however for the voluntary-interlock group they are still smaller than the percentages of participants having convictions in the control-voluntary group. Note that again, these results are not statistically significant (pvalues>0.05).

Figure 3-7: Percentages of participants with and without crashes after their inclusion date for control groups and after removal of interlock device for interlock groups. Top: all crashes, bottom: alcohol-related crashes

crashes_after	group: mand-inter	vol-inter	vol-contr	Total
do not have	91.30	90.85	90.57	90.91
have	8.70	9.15	9.43	9.09
Total	100.00	100.00	100.00	100.00

Pearson chi2(2) = 0.1074 Pr = 0.948

alc_crashes_after	group: mand-inte	vol-inter	vol-contr	Total
do not have	98.14	99.37	98.43	98.64
have	1.86	0.63	1.57	1.36
Total	100.00	100.00	100.00	100.00

Pearson chi2(2) = 1.9734 Pr = 0.373
Fisher's exact = 0.464

3.3 Conclusions

The descriptive statistics show that with respect to gender there are no significant differences between the respective experimental and control groups. However, there are some differences with respect to the distribution of age and condition 37. Most important are the comparisons in terms of alcohol-related convictions and crashes during the tracking period. The group with the smallest percentage of participants having alcohol-related convictions and crashes during the tracking period is the interlock-voluntary group versus the control-voluntary group which is the one with the largest percentage of participants having alcohol-related convictions and crashes (although the results for crashes are not statistically significant). These same results hold true after the interlock device was removed from the vehicle and although the percentage of participants in the voluntary-interlock with alcohol-related convictions and crashes increased, these percentages are still smaller than for those in the voluntary-control group. These results may suggest a positive impact of the alcohol ignition interlock program in reducing the alcohol-related convictions that may last even after the device is removed from the vehicle. This was further examined in detail in the following sections.

4. SURVIVAL ANALYSIS DURING AND AFTER THE PROGRAM

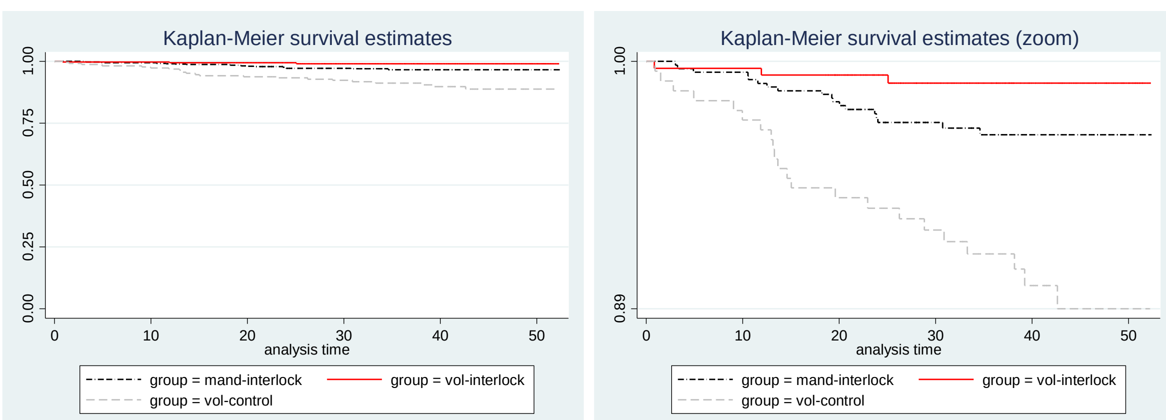
Survival analysis is used to compare the different study groups in terms of time to re-offend or crash during the tracking period based on their convictions and crash records related to alcohol (see appendix A for a description of the offence codes considered). The first subsection presents results with respect to convictions and the second with respect to crashes.

The analyses in this section consider time to re-offend and crash since inclusion in the study until the end of the study (June 2014). For the interlock groups the inclusion in the study was defined as the date the interlock device was installed. Therefore these analyses examine the impact of the program from the time the interlock device was installed e until the end of the study, whether the device was still installed or not. In the next section we present similar analyses but examining the impact of the program after the interlock device had been removed from the vehicle until the end of the study.

4.1 Convictions

The figure with the Kaplan-Meier survival estimates shows that the survival pattern is very different for the voluntary-control group in comparison to the other two interlock groups. Since the data for the mandatory-control group are very limited this group is not included in the analyses. The analysis time on the horizontal axis is measured in units of months and it represents the tracking period for each participant (since their inclusion date in the study until the end of June 2014).

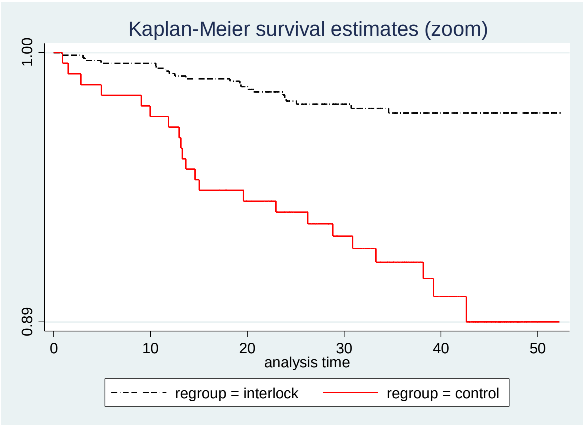
Figure 4-1: Kaplan Meier survival estimates for convictions for all groups



In the long term the probability of not having a new alcohol related conviction for the interlock groups is higher than for the voluntary-control, the ones who choose not to participate. In other words, it takes longer for the interlock groups to have an alcohol-related conviction than for the voluntary-control group. The figure also shows that the survival pattern is not very different for both interlock groups, in general the voluntary-interlock group has a more optimistic survival pattern (it takes longer before being convictedd) than the mandatory-interlock.

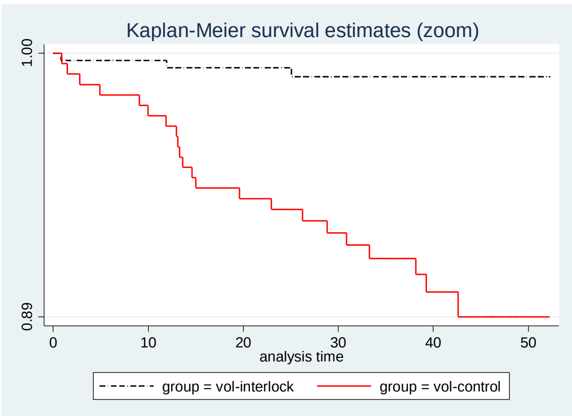
The descriptive analyses showed that in general both interlock groups are similar; therefore they can be combined together and compared to the control-voluntary group. When combining together both interlock groups and compare this with the voluntary-control group, the graph with the Kaplan-Meier survival estimates (see below) shows a clear distinction between the interlock and the control groups, in the same direction as before. The interlock group has a more optimistic survival pattern in relation to convictions than the control group.

Figure 4-2: Kaplan Meier survival estimates for convictions for interlock versus control



However, the log-rank test of equality for the survival functions rejects the null hypothesis that the survival functions of both interlock groups are the same (p-value=0.04). The test rejects the null hypothesis for the three groups overall and for the comparison of the combined interlock groups with the voluntary-control group (p-values<0.00005). The log-rank test also rejects the null hypothesis that the survival functions of both voluntary groups are equal (p-value<0.00005).

Figure 4-3: Kaplan-Meier survival estimates for convictions for the voluntary groups



Therefore in the following analyses, two types of comparisons are presented. First the comparison of the combined interlock groups with the control-voluntary and then the comparison of only the interlock-voluntary with the control-voluntary.

The log-log plot of the survival functions displays lines that are roughly parallel, suggesting that the proportional-hazards assumption is not violated and Cox regression is appropriate to estimate the association of each group to the alcohol-related conviction hazard rate. Furthermore, the tests of proportional-hazards assumption using Schoenfeld residuals did not find evidence to reject the assumption of proportional hazards (p-values>0.05).

Figure 4-4: Log-log plot of the survival functions for convictions

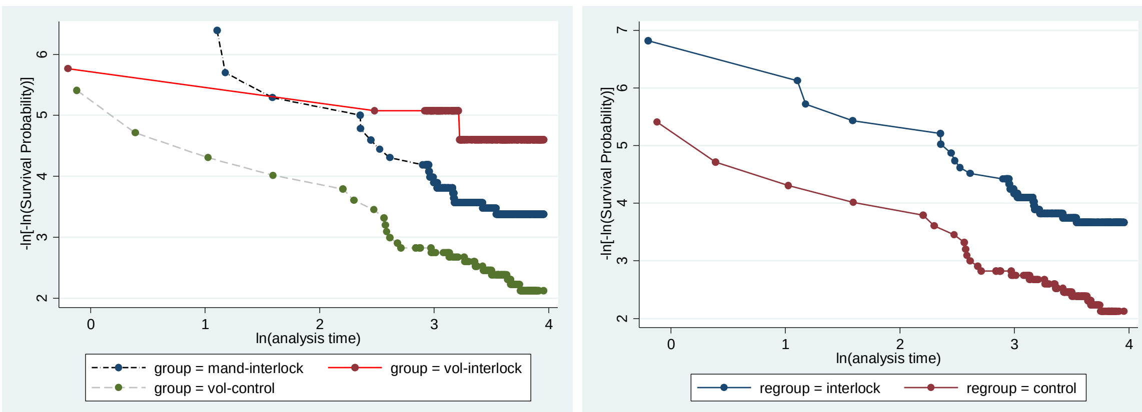


Figure 4-5: Test of proportional-hazards assumption for convictions

group	rho	chi2	df	Prob>chi2
1b.mand-int	.	.	1	.
2.vol-int	-0.09906	0.42	1	0.5162
3.vol-cont	0.02085	0.02	1	0.8922
global test		0.53	2	0.7670

regroup	rho	chi2	df	Prob>chi2
1b.interlock	.	.	1	.
2.control	0.05198	0.11	1	0.7357
global test		0.11	1	0.7357

The results from the Cox regression comparing the three groups show that the hazard ratio for the voluntary-control group is statistically significant (p-value<0.05) when compared to the hazard for the mandatory-interlock group. The hazard ratio of the voluntary interlock group compared to the mandatory interlock group is not statistically significant (p-value=0.058).

```
Cox regression -- Breslow method for ties
No. of subjects =          1141                     Number of obs   =       1157
No. of failures =            43
Time at risk    =   40731.11475
                                                    LR chi2(2)      =      24.60
Log likelihood  =    -283.77576                     Prob > chi2     =     0.0000
--------------------------------------------------------------------------------
           _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
        group |
vol-interlock |   .3071933   .1915702    -1.89   0.058     .0904878    1.042878
  vol-control |   3.099988   .9868523     3.55   0.000      1.66107    5.785383
--------------------------------------------------------------------------------
```

When combining together both interlock groups and comparing it with the voluntary-control group, the results from the Cox regression show that the hazard rate for the control group is 4.1times larger than the hazard rate for the interlock groups. This result is statistically significant (p-value<0.001).

```
Cox regression -- Breslow method for ties
No. of subjects =          1141                     Number of obs   =       1157
No. of failures =            43
Time at risk    =   40731.11475
                                                    LR chi2(1)      =      19.96
Log likelihood  =    -286.09983                     Prob > chi2     =     0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     regroup |
    control  |   4.099129   1.252673     4.62   0.000     2.252001      7.4613
------------------------------------------------------------------------------
```

When controlling for other possible cofactors, the results from the Cox regression show that age, gender and condition 37 are not statistically significant (p-value>0.05) to determine the hazard rate of the participants. However, note that according to this model the difference between interlock and control groups is more pronounced (4.6 hazard ratio versus 4.1 in the previous model) and it is still significant.

```
Cox regression -- Breslow method for ties
No. of subjects =          1137                     Number of obs   =       1153
No. of failures =            43
Time at risk    =   40552.06557
                                                    LR chi2(7)      =      21.77
Log likelihood  =    -285.01107                     Prob > chi2     =     0.0028
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     regroup |
    control  |    4.61962   1.588413     4.45   0.000     2.354662     9.06325
      agecat |
      25-34  |   1.224524   .6243911     0.40   0.691     .4507497    3.326591
      35-44  |   1.187941   .6484748     0.32   0.752     .4075115    3.462979
      45-64  |   1.201273   .6119847     0.36   0.719     .4425878    3.260496
  65 & over  |   1.300006   1.080789     0.32   0.752     .2548474    6.631481
      gender |
          M  |   1.637596   .8658722     0.93   0.351     .5809485    4.616109
      cond37 |   1.092357   .3757949     0.26   0.797     .5565864    2.143861
```

A more fair comparison is exclusively between both voluntary groups (because the voluntary-interlock group and the voluntary-control group are basically the same except for the fact that one group participates in the interlock program and the other did not participate in the program).

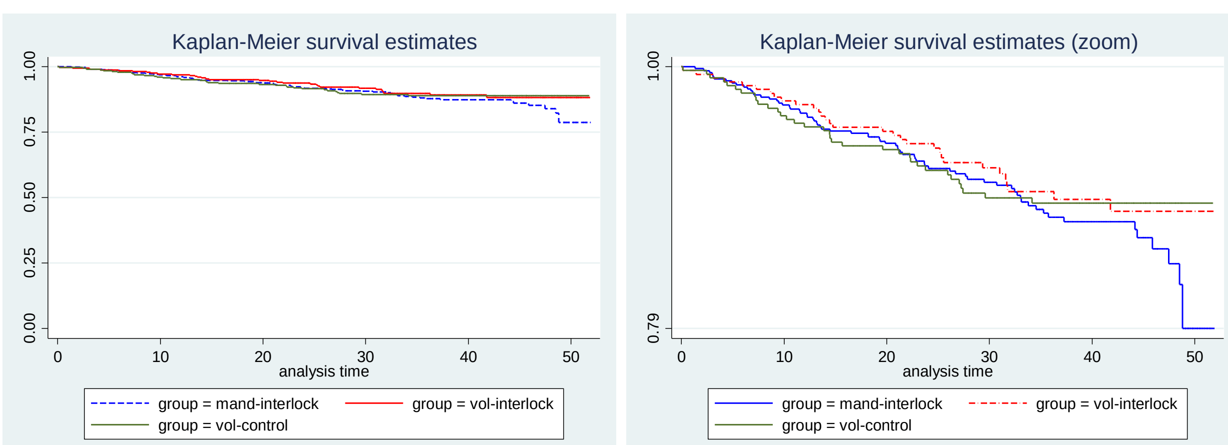
The Cox regression below comparing both voluntary groups, the voluntary-interlock and the voluntary-control, shows an even larger hazard ratio than the previous model (10.5 versus 4.6) which is still significant and again other factors like age, gender and condition 37, are not statistically significant (p-values>0.05).

```
Cox regression -- no ties
No. of subjects =          540                   Number of obs   =        549
No. of failures =           25
Time at risk    =  19619.31148
                                                 LR chi2(7)      =      26.56
Log likelihood  =   -139.54423                   Prob > chi2     =     0.0004
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |
 vol-control  |   10.45375   6.581506     3.73   0.000     3.043476    35.90661
      agecat |
      25-34  |   1.637528   .9162911     0.88   0.378     .5468803    4.903261
      35-44  |   1.118297    .757588     0.17   0.869     .2964233    4.218926
      45-64  |   1.046027   .6465986     0.07   0.942     .3114407    3.513263
  65 & over  |   1.265357    1.40145     0.21   0.832      .144365    11.09083
      gender |
          M  |   2.502066   1.853109     1.24   0.216     .5859654     10.6838
      cond37 |   1.448874   .6874405     0.78   0.435     .5716954    3.671946
```

4.2 Crashes

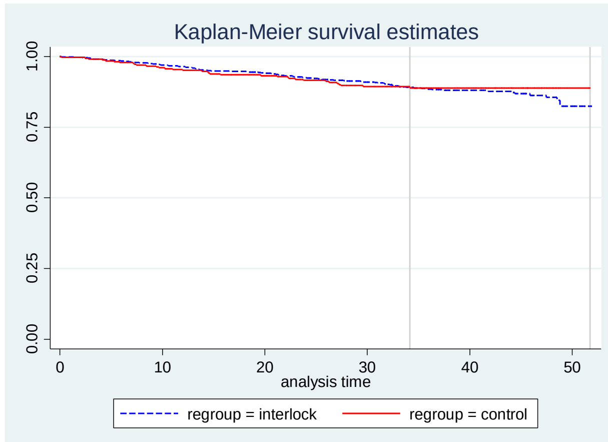
The figure with the Kaplan-Meier survival estimates shows that the survival pattern is not very different for the three groups. However, there seems to be some small differences. Comparing the interlock groups, in the long-term the voluntary -interlock group has a slightly more optimistic survival pattern (longer time to crash) than the mandatory-interlock. On the other hand, for the voluntary-control group, the survival function starts with a less optimistic pattern, below the survival function for the interlock groups. However, in the long-term it shows a cross-effect, surpassing the survival function of the voluntary-interlock group around time=35 months and of the mandatory-control around time=45 months.

Figure 4-6: Kaplan Meier survival estimates for crashes for all groups



The curve for the control group is completely flat from 34.2 to 51.7 analysis time since during that period none of the non-interlock participants have had any crashes.

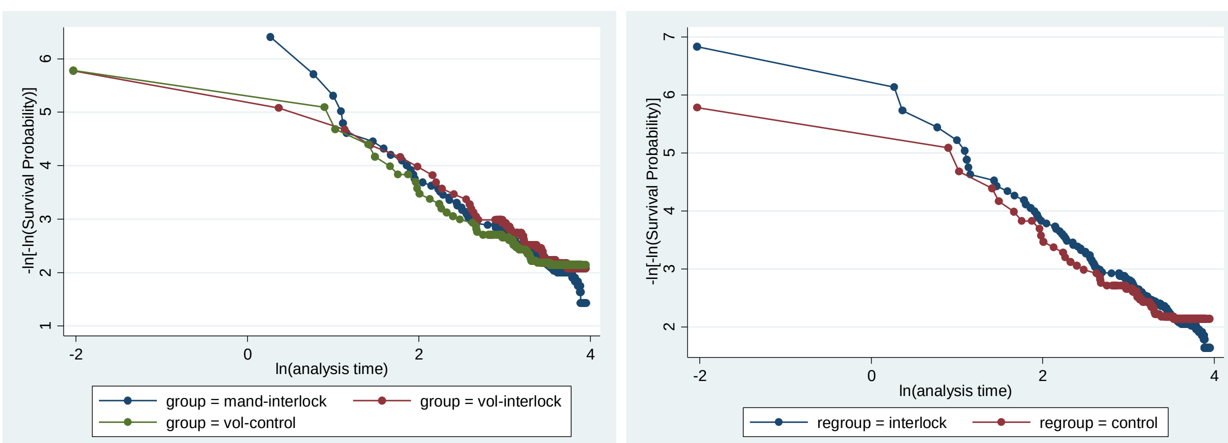
Figure 4 -7: Kaplan Meier survival estimates for crashes for interlock versus control



The log-rank test of equality for the survival functions does not reject the null hypothesis that the survival functions of the three groups are the same (p-value=0.5). A similar result is observed between the combined interlock groups and the voluntary control and also in the comparison of both interlock groups (p-value=0.68 and 0.27 respectively).

The log-log plot of the survival functions displays lines that do not seem parallel, suggesting that the proportional-hazards assumption may have been violated and Cox regression results should be interpreted with caution.

Figure 4-8: Log-log plot of the survival functions for convictions



The tests of proportional-hazards assumption using Schoenfeld residuals reject the assumption of proportional hazards, particularly between the combined interlock group and the control group (p-value < 0.05). This supports the idea that Cox regression results should be interpreted with caution.

Figure 4-9: Test of proportional-hazards assumption for crashes

```
       group   |        rho              chi2        df       Prob>chi2

    1b.mand-int |          .                 .         1               .
    2.vol-int   |     -0.08855            1.08         1          0.2979
    3.control   |     -0.19391            5.18         1          0.0228
    global test |                         5.31         2          0.0703
    ----------------------------------------------------------------
        regroup |        rho              chi2        df       Prob>chi2
    ------------+---------------------------------------------------
    1b.interlock|          .                 .         1               .
    2.control   |     -0.17426            4.20         1          0.0404
    global test |                         4.20         1          0.0404
```

The results from the different Cox regressions show that the hazard ratio for crashes between the different groups are not statistically significant (p-value>0.05).

```
Cox regression -- Breslow method for ties
No. of subjects =         1255                    Number of obs   =      1283
No. of failures =          137
Time at risk    =  42877.40984
                                                  LR chi2(2)      =      1.39
Log likelihood  =   -940.01219                    Prob > chi2     =    0.4980
--------------------------------------------------------------------------------
           _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+-----------------------------------------------------------------
         group |
 vol-interlock |   .7913795   .1700623   -1.09   0.276     .5193584    1.205876
   vol-control |   .8537507   .1779842   -0.76   0.448     .5673855    1.284647

Cox regression -- Breslow method for ties
No. of subjects =         1255                    Number of obs   =      1283
No. of failures =          137
Time at risk    =  42877.40984
                                                  LR chi2(1)      =      0.17
Log likelihood  =   -940.62254                    Prob > chi2     =    0.6770
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     regroup |
     control |   .9212251   .1826527   -0.41   0.679     .6245925    1.358735
```

When controlling for other possible cofactors, the results from the Cox regression show that age, gender and condition 37, are not statistically significant (p-value>0.05) to determine the hazard rate of the participants.

```
Cox regression -- Breslow method for ties
No. of subjects =         1137                    Number of obs   =      1165
No. of failures =          137
Time at risk    =  38586.22951
                                                  LR chi2(7)      =      8.58
Log likelihood  =   -922.44153                    Prob > chi2     =    0.2843
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     regroup |
     control |   1.233743   .2720877    0.95   0.341     .8007595    1.900848
      agecat |
      25-34  |   .9057242   .2650993   -0.34   0.735     .5103347    1.607448
      35-44  |   .5793878   .1871858   -1.69   0.091     .3075865    1.091369
      45-64  |   .7115642    .209342   -1.16   0.247     .3997521    1.266594
   65 & over |   .5402481   .2786351   -1.19   0.233     .1965996    1.484581
      gender |
          M  |   1.428719   .4194299    1.22   0.224     .8036352    2.540005
      cond37 |   .9935305   .1810849   -0.04   0.972     .6950848    1.420119
```

The Cox regression below comparing both voluntary groups, the voluntary-interlock and the voluntary-control, also shows no statistically significant results (p-values>0.05).

```
Cox regression -- Breslow method for ties
 No. of subjects =          540                    Number of obs   =       554
 No. of failures =           65
 Time at risk    =   18649.2459
                                                   LR chi2(7)      =      9.35
 Log likelihood  =   -394.11515                    Prob > chi2     =    0.2283
 ----------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
 -------------+--------------------------------------------------------------
        group |
 vol-control  |   1.367437   .3610393     1.19   0.236     .8150186    2.294282
       agecat |
       25-34  |   1.355024   .4895895     0.84   0.400     .6674147    2.751049
       35-44  |   .6344833   .2858394    -1.01   0.313     .2623902    1.534238
       45-64  |   .6937221   .2734809    -0.93   0.354     .3203486    1.502271
   65 & over  |   1.064963    .632172     0.11   0.916     .3327005    3.408912
              |
       gender |
           M  |   1.138353   .3975668     0.37   0.711     .5741124    2.257131
       cond37 |   .8681917   .2797348    -0.44   0.661     .4616949    1.632586
```

In case the proportional hazard assumption does not hold, better parametric models may be an alternative to the semi-parametric Cox model. The figure below shows the results from a flexible parametric model using the hazard scale with 4 degrees of freedom for the baseline hazard function. However, this model does not reveal any new significant information.

```
 Log likelihood = -549.53619                      Number of obs   =      1165
 ----------------------------------------------------------------------------
              |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
 -------------+--------------------------------------------------------------
 xb           |
        group |   1.046814   .1264708     0.38   0.705     .8260981      1.3265
          age |   .9909515    .006573    -1.37   0.171     .9781521    1.003918
       gender |   1.394053   .4085043     1.13   0.257     .7849621    2.475768
       cond37 |   .9706104   .1867477    -0.16   0.877     .6656896    1.415201
        _rcs1 |   1.681035   .0756749    11.54   0.000      1.53907    1.836095
        _rcs2 |   1.018728   .0450057     0.42   0.674     .9342299     1.11087
        _rcs3 |   1.027393   .0295302     0.94   0.347     .9711145    1.086932
        _rcs4 |   .9914319   .0166539    -0.51   0.608     .9593224    1.024616
        _cons |   .0798166   .0554338    -3.64   0.000     .0204608    .3113617
```

4.3 Conclusions

The influence of the interlock program was examined in terms of convictions and crashes using survival analysis during the entire study period, i.e., both when the interlock was installed and after it was removed. With respect to convictions, the participants in the interlock program are less prone to have alcohol-related convictions as the survival analysis revealed it takes longer for a conviction to occur compared to the voluntary-control participants. Although the voluntary-interlock group seems to be slightly less prone to have alcohol-related convictions in the long-term than the mandatory-interlock group, the differences are not statistically significant. In other words, the survival analyses support the notion that the interlock program is associated with a positive impact on reducing the risk for alcohol-related convictions, and there seems to be no difference in this respect between mandatory and voluntary participants. This overall finding stands even when comparing the voluntary-interlock group with the voluntary-control group only. Therefore, it can be concluded that the data suggest the interlock program has a very strong effect on the behaviour of all interlocked offenders. To illustrate, the hazard rate for being convicted is 10.5 times larger for offenders in the voluntary control group compared to offenders in the voluntary interlock group.

With respect to crashes there seems to be no statistically significant differences between the participants of the interlock and voluntary-control groups.

It warrants mentioning that since the amount of data in the analyses is not very large, similar analyses were conducted using extra information for the interlock groups. The extra information consisted of including participants in the interlock program who enrolled before the intake period (from November 2008 to February 2010). While it can be argued that this unbalanced design would likely bias the results, overall, the findings were the same.

5. SURVIVAL ANALYSIS AFTER THE PROGRAM

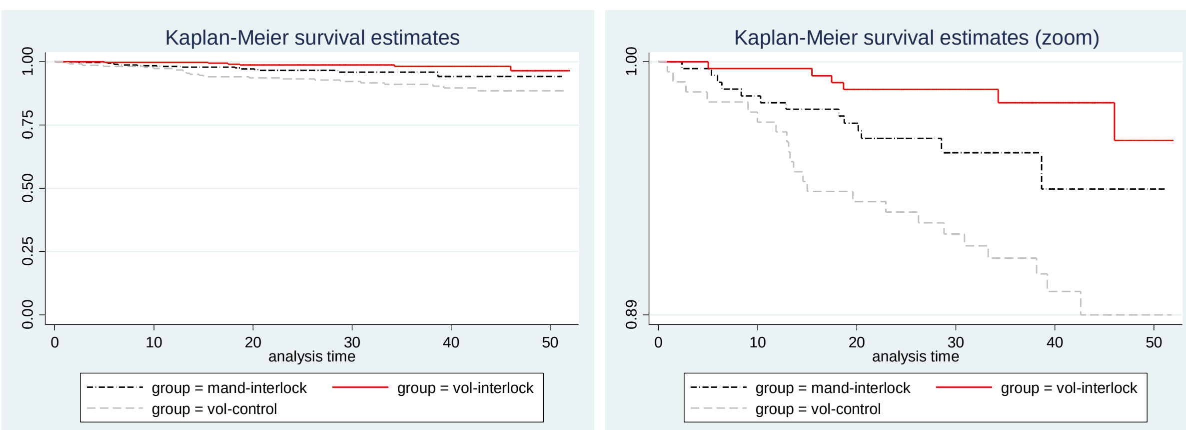
The survival analyses in the previous section examine the impact of the program since the installation of the device until the end of the study, whether the interlock device was still installed or not. In this section we present survival analyses examining the impact of the program after the interlock device had been removed from the vehicle. For these analyses, only those who had an interlock device removed in the period March 2010-December 2012 are considered in the interlock experimental groups.

As in the previous section, survival analysis is used to compare the different study groups in terms of time to re-offend during the tracking period based on their convictions and crash records (see appendix A for a description of the offence codes considered). The first subsection presents results with respect to convictions and the second with respect to crashes.

5.1 Convictions

The figure with the Kaplan-Meier survival estimates shows that the survival pattern is very different for the voluntary-control group in comparison to the other two interlock groups. The analysis time on the horizontal axis is measured in units of months and it represents the tracking period for each participant. Since the data for the mandatory-control group are very limited this group is not included in the analyses.

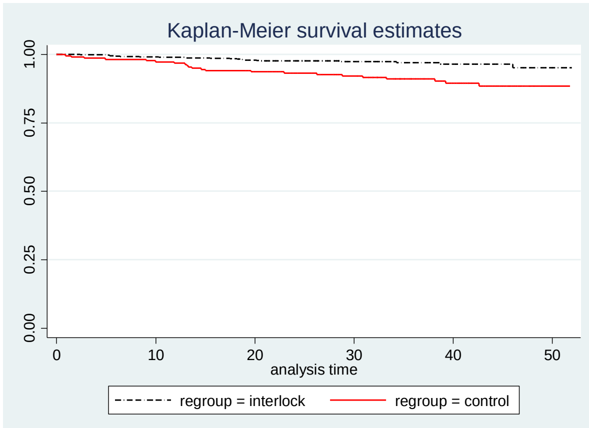
Figure 5-1: Kaplan Meier survival estimates for convictions for all groups



In the long term the probability of not having a new alcohol related conviction for the interlock groups is higher than for the voluntary-control, the ones who choose not to participate. In other words, it takes longer for the interlock groups to have an alcohol-related conviction than for the voluntary-control group. The figure also shows that the survival pattern is not very different for both interlock groups, in general the voluntary-interlock group has a more optimistic survival pattern (it takes longer before being convicted) than the mandatory-interlock.

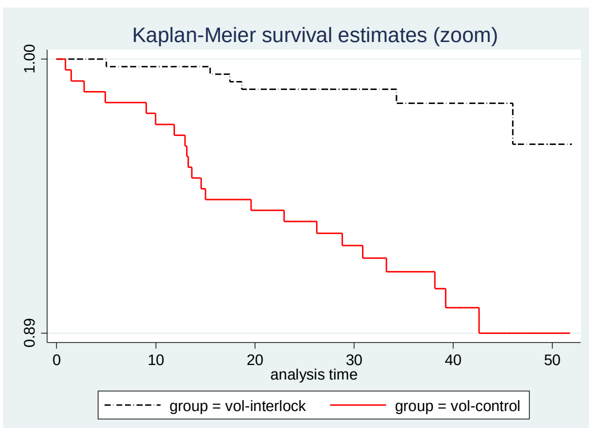
The descriptive analyses showed that in general both interlock groups are similar; therefore they can be combined together and compared to the control-voluntary group. When combining together both interlock groups and compare this with the voluntary-control group, the graph with the Kaplan-Meier survival estimates (see below) shows a clear distinction between the interlock and the control groups, in the same direction as before. The interlock group has a more optimistic survival pattern in relation to convictions than the control group.

Figure 5-2: Kaplan Meier survival estimates for convictions for interlock versus control



However, the log-rank test of equality for the survival functions rejects the null hypothesis that the survival functions of both interlock groups are the same (p-value=0.05). The test rejects the null hypothesis for the three groups overall and for the comparison of the combined interlock groups with the voluntary mandatory (p-values<0.00005). The log-rank test also rejects the null hypothesis that the survival functions of both voluntary groups are equal (p-value<0.05).

Figure 5-3: Kaplan-Meier survival estimates for convictions for the voluntary groups



Therefore in the following analyses, two types of comparisons are presented. First the comparison of the combined interlock groups with the control-voluntary and then the comparison of only the interlock-mandatory with the control-voluntary.

The log-log plot of the survival functions displays lines that are roughly parallel, suggesting that the proportional-hazards assumption is not violated and Cox regression is appropriate to estimate the association of each group to the alcohol-related conviction hazard rate. Furthermore, the tests of proportional-hazards assumption using Schoenfeld residuals did not find evidence to reject the assumption of proportional hazards (p-values>0.05).

Figure 5-4: Log-log plot of the survival functions for convictions

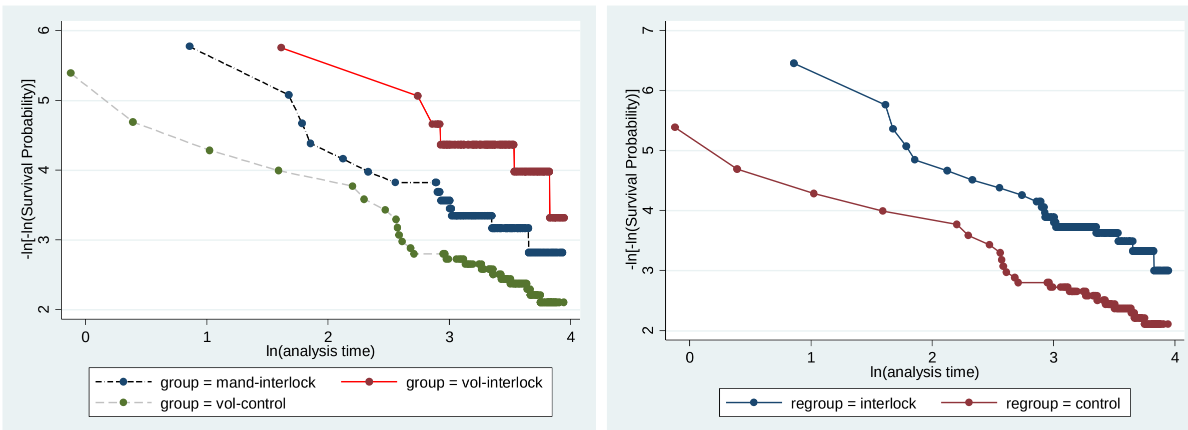


Figure 5-5: Test of proportional-hazards assumption for convictions

group	rho	chi2	df	Prob>chi2
1b.mand-int	.	.	1	.
2.vol-int	0.11878	0.57	1	0.4513
3.vol-cont	-0.04233	0.07	1	0.7856
global test		1.04	2	0.5944

regroup	rho	chi2	df	Prob>chi2
1b.interlock	.	.	1	.
2.control	-0.07334	0.22	1	0.6385
global test		0.22	1	0.6385

The results from the Cox regression comparing the three groups show that the hazard ratios for the voluntary groups are not statistically significant at the 5% level (p-values>0.05) when compared to the hazard for the mandatory-interlock group, but they are at the 10% level (p-values<0.1).

```
Cox regression -- no ties
No. of subjects =          858                  Number of obs   =        878
No. of failures =           41
Time at risk    =        28834
                                                LR chi2(2)      =      15.36
Log likelihood  =   -258.83552                  Prob > chi2     =     0.0005
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
group						
vol-interlock	.398017	.1972997	-1.86	0.063	.1506452	1.051593
vol-control	1.992258	.7086574	1.94	0.053	.9921299	4.000578

When combining together both interlock groups and comparing it with the voluntary-control group, the results from the Cox regression show that the hazard rate for the control group is 3 times larger than the hazard rate for the interlock groups. This result is statistically significant (p-value=0.001).

```
Cox regression -- no ties
No. of subjects =          858                     Number of obs   =       878
No. of failures =           41
Time at risk    =        28834
                                                   LR chi2(1)      =     11.62
Log likelihood  =    -260.7082                     Prob > chi2     =    0.0007
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     regroup |
     control |   2.964767   .9354525     3.44   0.001     1.597412    5.502554
```

When controlling for other possible cofactors, the results from the Cox regression show that age, gender and condition 37 are not statistically significant (p-value>0.05) to determine the hazard rate of the participants. However, note that according to this model the difference between interlock and control groups is more pronounced (3.6 hazard ratio versus 3 in the previous model) and it is still significant.

```
Cox regression -- no ties
No. of subjects =          854                     Number of obs   =       874
No. of failures =           41
Time at risk    =  28656.78689
                                                   LR chi2(7)      =     23.44
Log likelihood  =    -254.5364                     Prob > chi2     =    0.0014
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     regroup |
     control |   3.555764   1.218692     3.70   0.000      1.81633    6.960991
      agecat |
      25-34  |   2.046418   1.072059     1.37   0.172     .7329503    5.713657
      35-44  |    .659065   .4519337    -0.61   0.543     .1718869    2.527049
      45-64  |   1.523604   .8163399     0.79   0.432     .5330919    4.354536
   65 & over |   .7517988   .8345194    -0.26   0.797     .0853576    6.621573
      gender |
          M  |   2.222629   1.344102     1.32   0.187     .6793842    7.271408
      cond37 |   1.810988   .6059735     1.77   0.076       .93993    3.489278
```

As mentioned previously, a more fair comparison is exclusively between both voluntary groups. The Cox regression below comparing both voluntary groups, the voluntary-interlock and the voluntary-control, shows an even larger hazard ratio than the previous model (5.4 versus 3.6) which is still significant and again other factors like age, gender and condition 37, are not statistically significant (p-values>0.05).

```
Cox regression -- no ties
No. of subjects =          532                     Number of obs   =       543
No. of failures =           28
Time at risk    =  19207.14754
                                                   LR chi2(7)      =     21.64
Log likelihood  =    -158.3741                     Prob > chi2     =    0.0029
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |
vol-control  |   5.413332   2.603622     3.51   0.000     2.108952    13.89513
      agecat |
      25-34  |   2.242189   1.330698     1.36   0.174     .7006477    7.175377
      35-44  |   1.267883   .9072686     0.33   0.740     .3118737     5.15442
      45-64  |   1.651246   1.022369     0.81   0.418     .4906706    5.556911
   65 & over |   1.534309   1.740861     0.38   0.706     .1660048    14.18094
      gender |
          M  |   2.880981   2.125738     1.43   0.152     .6783901    12.23492
      cond37 |   1.548307   .6818545     0.99   0.321     .6531286    3.670416
```

5.2 Crashes

The figure with the Kaplan-Meier survival estimates shows that the survival pattern is not very different for the three groups.

Figure 5-6: Kaplan Meier survival estimates for crashes for all groups

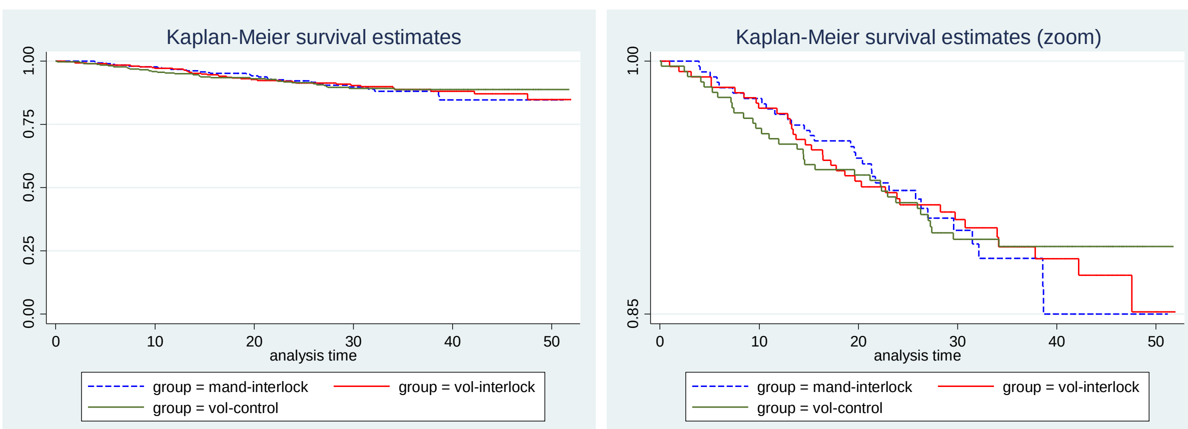
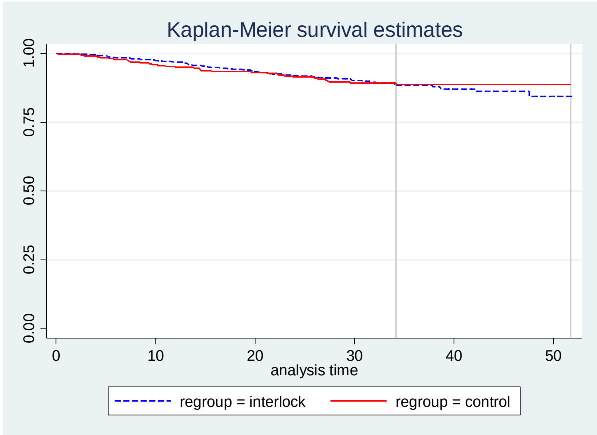


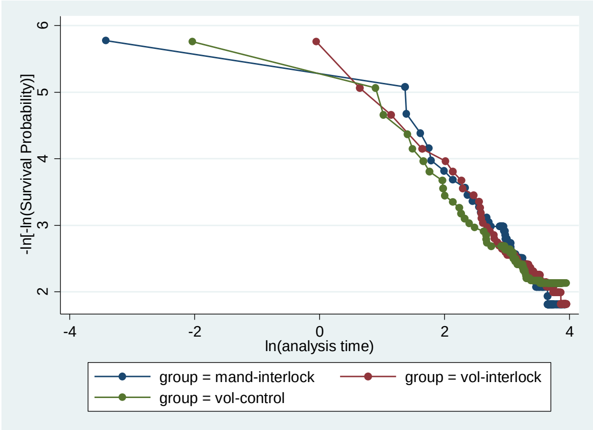
Figure 5-7:: Kaplan Meier survival estimates for crashes for interlock versus control

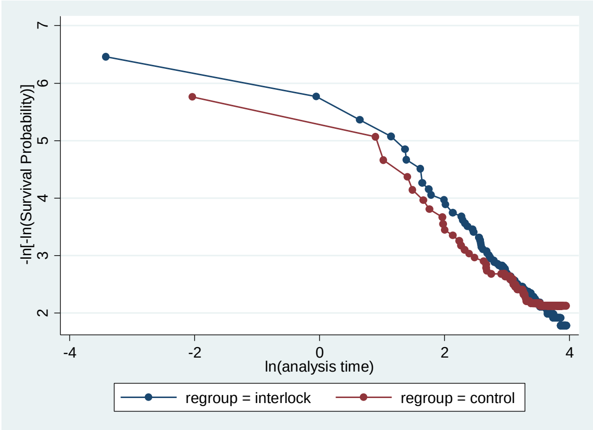


The log-rank test of equality for the survival functions does not reject the null hypothesis that the survival functions of the three groups are the same (p-value=0.9). A similar result is observed between the combined interlock groups and the voluntary control and also in the comparison of both interlock groups (p-value=0.7 and 0.8 respectively).

The log-log plot of the survival functions displays lines that do not seem parallel, suggesting that the proportional-hazards assumption may have been violated and Cox regression results should be interpreted with caution.

Figure 5-8: Log-log plot of the survival functions for convictions





The tests of proportional-hazards assumption using Schoenfeld residuals reject the assumption of proportional hazards, particularly between the combined interlock group and the control group (p-value < 0.05). This supports the idea that Cox regression results should be interpreted with caution.

Figure 5-9: Test of proportional-hazards assumption for crashes

```
      group  |      rho            chi2       df       Prob>chi2
-----------+--------------------------------------------------
1b.man-int  |        .               .         1             .
2.vol-int   |   -0.05157          0.27         1        0.6034
3.control   |   -0.20283          4.37         1        0.0366
-----------+--------------------------------------------------
global test |                     4.82         2        0.0897
--------------------------------------------------------------
--------------------------------------------------------------
    regroup  |      rho            chi2       df       Prob>chi2
-----------+--------------------------------------------------
1b.interlock|        .               .         1             .
2.control   |   -0.20753          4.51         1        0.0336
global test |                     4.51         1        0.0336
```

The results from the different Cox regressions show that the hazard ratio for crashes between the different groups are not statistically significant (p-value>0.05).

```
Cox regression -- Breslow method for ties
No. of subjects =          957                     Number of obs   =       978
No. of failures =          101
Time at risk    =  31413.31148
                                                   LR chi2(2)      =      0.17
Log likelihood   =   -672.01309                    Prob > chi2     =    0.9172
--------------------------------------------------------------------------------
           _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+-----------------------------------------------------------------
        group |
 vol-interlock |   .9684529   .2379345    -0.13   0.896      .598344    1.567495
  vol-control |    .904655   .2245904    -0.40   0.686     .5561128    1.471645
--------------------------------------------------------------------------------
Cox regression -- Breslow method for ties
No. of subjects =          957                     Number of obs   =       978
No. of failures =          101
Time at risk    =  31413.31148
                                                   LR chi2(1)      =      0.16
Log likelihood   =     -672.0216                   Prob > chi2     =    0.6931
           _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
      regroup |
      control |   .9202375   .1946537    -0.39   0.694     .6079238    1.392999
```

When controlling for other possible cofactors, the results from the Cox regression show that gender, age and condition 37, are not statistically significant (p-value>0.05) to determine the hazard rate of the participants.

```
Cox regression -- Breslow method for ties
No. of subjects =          854                     Number of obs   =       875
No. of failures =          101
Time at risk    =  27613.18033
                                                   LR chi2(7)      =     10.02
Log likelihood   =   -654.48657                    Prob > chi2     =    0.1876
--------------------------------------------------------------------------------
           _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
      regroup |
      control |   1.132933   .2644223     0.53   0.593     .7170291    1.790075
       agecat |
        25-34 |   .9462691   .2934099    -0.18   0.859     .5153264    1.737588
        35-44 |   .5002281   .1815069    -1.91   0.056     .2456467     1.01865
        45-64 |    .595976   .1895358    -1.63   0.104     .3195396     1.11156
    65 & over |   .4017387   .2564868    -1.43   0.153     .1149479    1.404062
       gender |
            M |   .9355316   .2729462    -0.23   0.819     .5280997    1.657299
       cond37 |   1.009864   .2249037     0.04   0.965     .6526711    1.562541
```

The Cox regression below comparing both voluntary groups, the voluntary-interlock and the voluntary-control, also shows no statistically significant results (p-values>0.05) in terms of groups, gender and condition 37. However, the age groups 35-44 and 45-64 have significantly smaller hazard rate than the younger baseline group 15-24 (hazard ratios 0.37 and 0.47 respectively).

```
Cox regression -- Breslow method for ties
No. of subjects =          532                     Number of obs   =       548
No. of failures =           69
Time at risk    =   18479.2459
                                                   LR chi2(7)      =     12.50
Log likelihood  =   -414.93429                     Prob > chi2     =    0.0852
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |
 vol-control |   1.075847   .2813234     0.28   0.780     .6444209    1.796103
             |
      agecat |
      25-34  |   .9940882   .3287502    -0.02   0.986     .5199098    1.900736
      35-44  |   .3712049   .1649875    -2.23   0.026     .1553405     .887039
      45-64  |   .4655123   .1687919    -2.11   0.035      .228713    .9474834
  65 & over  |   .5572245   .3603647    -0.90   0.366     .1568723    1.979311
      gender |
          M  |    .959178   .3093453    -0.13   0.897     .5097738    1.804766
      cond37 |   1.085054   .3352641     0.26   0.792     .5921665    1.988194
```

In case the proportional hazard assumption does not hold, better parametric models may be an alternative to the semi-parametric Cox model. The figure below shows the results from a flexible parametric model using the hazard scale with 4 degrees of freedom for the baseline hazard function. However, this model does not reveal any new significant information.

```
Log likelihood = -407.13241                       Number of obs   =        875
------------------------------------------------------------------------------
             |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
xb           |
       group |   1.088505   .1527162     0.60   0.546     .8268134    1.433023
         age |   .9842386   .0077047    -2.03   0.042     .9692529    .9994561
      gender |   .9299524   .2711547    -0.25   0.803     .5251315    1.646847
      cond37 |   1.017868   .2358363     0.08   0.939     .6463546    1.602919
       _rcs1 |   1.631121    .082492     9.67   0.000     1.477194    1.801087
       _rcs2 |   .9935087   .0430441    -0.15   0.881     .9126266    1.081559
       _rcs3 |   1.081339   .0334616     2.53   0.012     1.017704    1.148952
       _rcs4 |   1.009428   .0194542     0.49   0.626     .9720098    1.048287
       _cons |   .2024527   .1532072    -2.11   0.035      .045938    .8922263
```

5.3 Conclusions

The results in this section (survival time since the interlock device was removed until the end of the study) are similar to the results in the previous section (survival time since the interlock device was installed until the end of the study). With respect to convictions, the risk for having an alcohol-related conviction is significantly larger for the voluntary-control group compared to the risk for both interlock groups combined as well as compared to the risk of the voluntary-interlock group alone. With respect to crashes, in both cases there seems to be no statistically significant differences between the interlock and the control groups.

Although the differences between interlock and control groups in terms of risk for convictions are less pronounced after the device was removed from the vehicle (hazard ratios in this section are less pronounced than in the previous section), the interlock program still has a significant and positive effect. This suggests a positive effect of the program reducing the risk for alcohol related convictions that remains, albeit less pronounced, after exiting the program. The less pronounced effect of the interlock on the behaviour of interlocked offenders after removing the device is a logical finding and it is consistent with the literature (see Alcohol Interlock curriculum for practitioners http://aic.tirf.ca/section2/qa.php#q8).

6. TIME SERIES ANALYSIS

In this section the results from the time series analyses are described. A subsection is devoted to each time series (monthly counts of alcohol related convictions, convictions and crashes) explaining the different steps involved in the building of the model as well as the results from each final model (see appendix A for a description of the offence codes considered).

As explained in the methodology, using data only from the period preceding the implementation of the AIIP (i.e., January 1998 through to September 2008, inclusive), the structure of the experimental time series was investigated. Special attention was given to the overall pattern, outliers, stationarity, seasonality and variance of the data as required for the ARMA time series models. Once the final model for the pre-intervention data was selected, variables were entered simultaneously to test whether the impact of the implementation of AIIP was significant. Three different models were tested: a sudden permanent model, a gradual permanent model and a sudden temporary model. The three models were analyzed using dummy variables to identify the time when the impact, if any, took place. Thirteen different months were tested for this purpose: the month of the intervention (the implementation of AIIP), the first month following the intervention, the second month following the intervention and so on until the twelfth month following the intervention. In other words, for each of the three different models, 13 intervention months were studied, for a total of 39 models (3X13).

The first model, the sudden permanent model, was tested using dummy variables representing the start of the effect of the program implementation (IIPimpl: value 0 for all months before the start of the effect of the implementation and value 1 for all months at, or following, the start of the effect). The second model, the gradual permanent model, was tested using an interaction between this dummy variable for the start of the effect and the post-intervention trend (TIIPimpl: value 0 for all months before the start of the effect of the implementation and increasing values for all months at, or following, the start of the effect). The third model, the sudden temporary model, was tested using a pulse variable (value 1 for the start of the effect month and 0 for all other months).

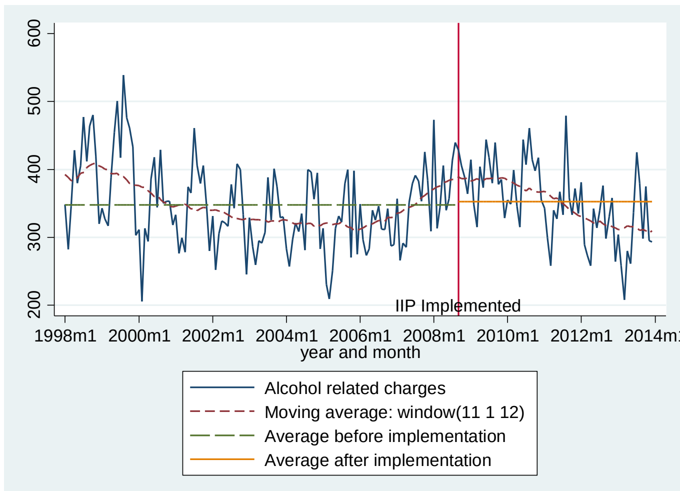
Other variables that were simultaneously entered in the model included:

- Population: Population estimates by quarters, aged 16 and over, 1998 to 2013 (Statistics Canada, 2014);
- Unemployment rate: Monthly percentage of adults aged 15 and over in the labour force that are unemployed, (Statistics Canada, 2014);
- Heavy drinking: Annual population aged 12 and over who reported having 5 or more drinks on one occasion, at least once a month in the past 12 months, 1998 to 2012 (before 2008 available information was biannual), (Statistics Canada, 2013); and,
- Alcohol sales: Average litres bought annually by adults aged 15 and older, 1998-2013 (Statistics Canada, 2014), Statistics Canada, CANSIM Table 183–0019 (per capita consumption estimates determined using population aged 15 years and over).

6.1 Charges

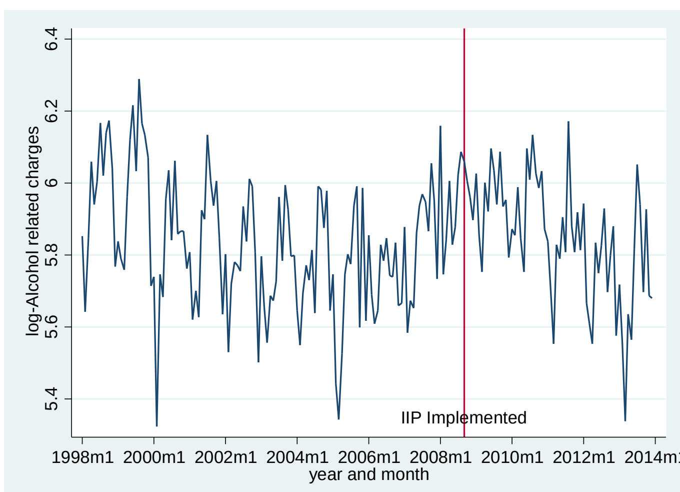
The figure below shows the time series of monthly counts of alcohol related charges. The vertical line in this figure represents the month when AIIP was implemented (September 2008). The moving average is also included in which each observation is an average of 24 nearby observations in the original series (11months backward, the actual observation and 12 months forward). The figure also shows the mean over the pre-implementation period (from January 1998 to August 2008) and the mean in the post-implementation period (from September 2008 to December 2013).

Figure 6-1: Monthly counts of alcohol related charges



To mitigate the impact of outliers and changing variance over time, these monthly counts were log-transformed. The result of this log-transformed time series is shown below.

Figure 6-2: Log-transformed series of monthly counts of alcohol related charges



The figure below shows the time series for non-alcohol related charges used as a control group. In the next figure the log-transformed series of both alcohol and non-alcohol related charges are compared.

Figure 6-3: Monthly counts of non-alcohol related charges

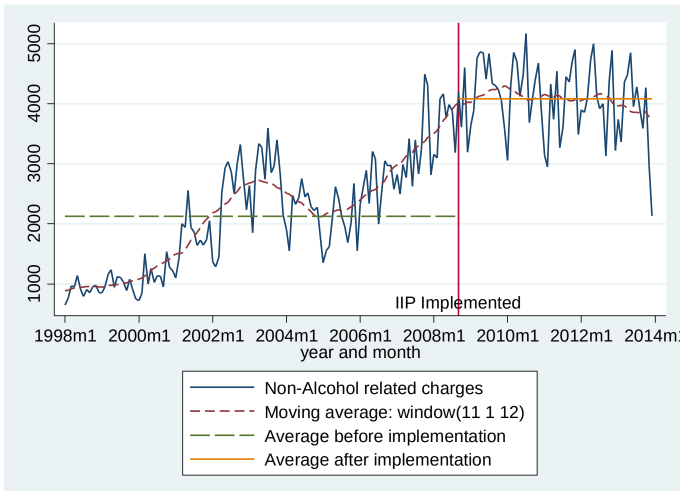
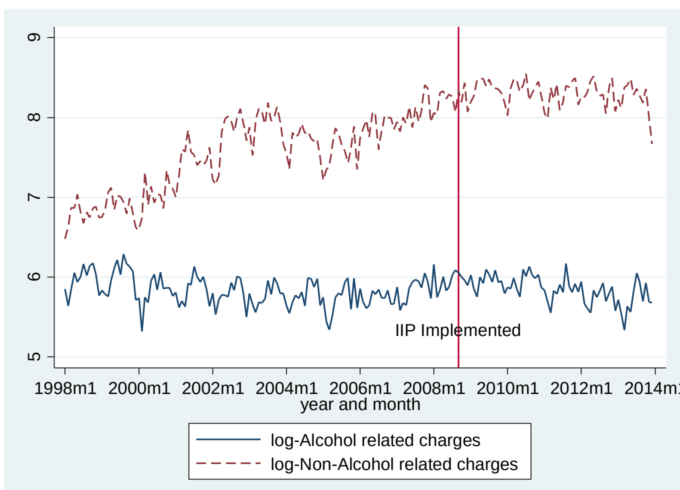


Figure 6-4: Log-transformed series of monthly counts of alcohol and non-alcohol related charges



Model selection for pre-intervention data

When looking at the pre-intervention series (from January 1998 to August 2008), there does not appear to be an increasing trend but there does appear to be seasonal variation in the time series, suggesting seasonal differencing may be required to make this series stationary.

A regression model was used to formally test the presence of a trend in the pre-intervention series. Figure 6-5 contains the results from this model. As can be seen, the trend is significant, (coef.: -0.001; s.e.: 0.0004; p:0.013), suggesting that local differencing might be necessary.

Figure 6-5: Regression model of a trend for the pre-intervention series

```
      Source |       SS       df       MS              Number of obs =     128
-------------+------------------------------           F(  1,   126) =    6.30
       Model |  .199724308     1  .199724308           Prob > F      =  0.0133
    Residual |  3.99179408   126  .031680905           R-squared     =  0.0476
-------------+------------------------------           Adj R-squared =  0.0401
       Total |  4.19151839   127  .033004082           Root MSE      =  .17799
------------------------------------------------------------------------------
lnCharge_alc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       trend |  -.0010691   .0004258    -2.51   0.013    -.0019117   -.0002265
       _cons |   5.903601     .03165   186.53   0.000     5.840967    5.966235
```

The time series of log-transformed counts were also standardized by creating Z-scores to identify outliers. An observation with a Z-score greater than 2.5 was considered an outlier. Three such outliers were identified for August 1999, February 2000 and March 2005 (Z-scores: 2.504, 2.8 and 2.7). These outliers are tagged in the estimated models using dummy variables.

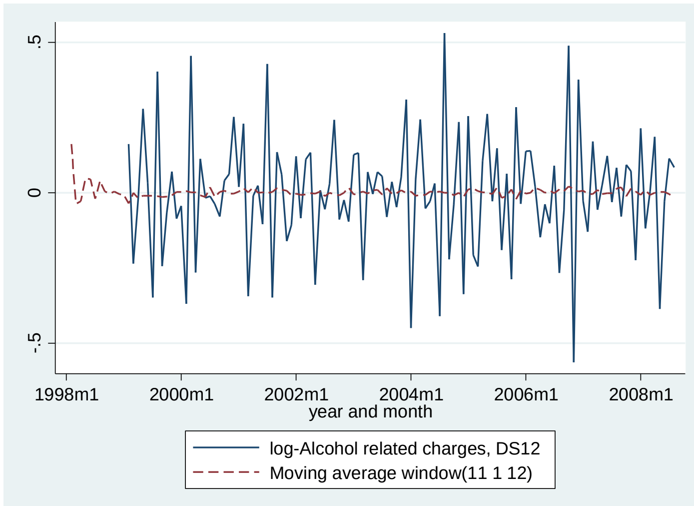
A correlogram was used to further investigate the need for seasonal differencing. The results for the log-transformed series are displayed in Figure 6-6. This correlogram confirms that there is considerable seasonal variation in the data, suggesting seasonal differencing is required to make the time series stationary.

Figure 6-6: Correlogram of the time series

```
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.5058   0.5137   33.523  0.0000          |----              |----
2        0.4357   0.2493   58.589  0.0000          |---               |-
3        0.2287  -0.0929    65.55  0.0000          |-                 |
4       -0.0484  -0.3312   65.864  0.0000          |                --|
5       -0.1337  -0.0859    68.28  0.0000         -|                  |
6       -0.1948   0.0572   73.454  0.0000         -|                  |
7       -0.1552   0.0904   76.768  0.0000         -|                  |
8       -0.0659   0.0795   77.369  0.0000          |                  |
9        0.1043   0.1826   78.891  0.0000          |                  |-
10       0.2856   0.2444    90.39  0.0000          |--                |-
11       0.3604   0.1121   108.86  0.0000          |--                |
12       0.5399   0.3414   150.68  0.0000          |----              |--
13       0.4045   0.0060   174.35  0.0000          |---               |
14       0.3320  -0.0245   190.44  0.0000          |--                |
15       0.0632  -0.2650   191.03  0.0000          |                --|
16      -0.0984  -0.0846   192.47  0.0000          |                  |
17      -0.2602  -0.1591   202.62  0.0000        --|                 -|
18      -0.2838   0.0043   214.81  0.0000        --|                  |
19      -0.2520  -0.0041    224.5  0.0000        --|                  |
20      -0.1383   0.0639   227.45  0.0000         -|                  |
21       0.0087   0.0150   227.46  0.0000          |                  |
22       0.1280  -0.0403   230.03  0.0000          |-                 |
23       0.2989   0.1975   244.19  0.0000          |--                |-
24       0.3624   0.1641    265.2  0.0000          |--                |-
25       0.2859  -0.0363   278.41  0.0000          |--                |
26       0.2946   0.0911   292.56  0.0000          |--                |
27      -0.0122  -0.2273   292.59  0.0000          |                 -|
28      -0.0707  -0.0673   293.42  0.0000          |                  |
29      -0.2140  -0.0350   301.12  0.0000         -|                  |
30      -0.3210  -0.1631   318.61  0.0000        --|                 -|
```

The figure below shows the local and seasonal differenced log-transformed time series. The plot shows that this series now seems to be stationary around a constant mean after the different transformations.

Figure 6-7: Local and seasonal differenced log-transformed time series and its moving average



Two unit root tests (Phillips-Perron and Dickey-Fuller) rejected the null hypothesis of a unit root (p-value<0.05), thus confirming the stationarity of the local and seasonal differenced series.

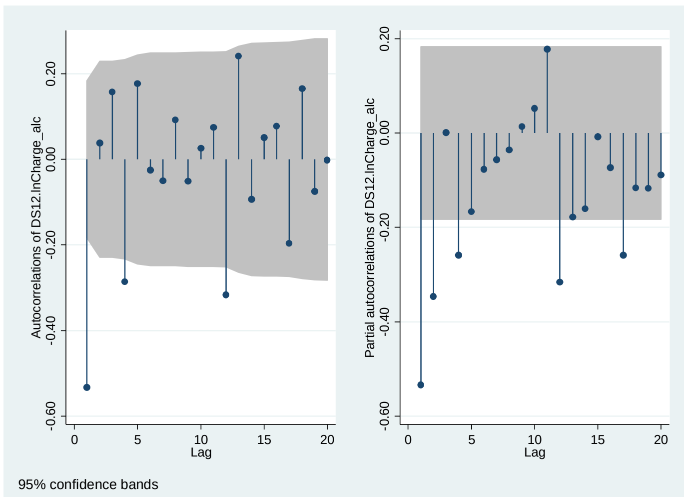
Figure 6-8: Two unit root tests: Phillips-Perron and Dickey-Fuller

```
 Phillips-Perron test for unit root                Number of obs   =       114
                                                  Newey-West lags =         4
                              ---------- Interpolated Dickey-Fuller ---------
                 Test         1% Critical       5% Critical      10% Critical
              Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(rho)         -151.473          -19.847           -13.728           -11.019
 Z(t)            -25.572           -3.505            -2.889            -2.579
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
Dickey-Fuller test for unit root                  Number of obs   =       114
                              ---------- Interpolated Dickey-Fuller ---------
                 Test         1% Critical       5% Critical      10% Critical
              Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -19.244           -3.505            -2.889            -2.579
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

Once a time series has been rendered stationary by differencing, the choice of parameters for the ARMA time series model may be made by examining two time-domain constructs: the autocorrelation function (ACF) and the partial autocorrelation function (PACF). In Figure 6-9, the autocorrelations and partial correlations of the seasonally and locally differenced pre-intervention time series are displayed. The autocorrelations can be used to better understand the structure of the series and to test whether the series is stationary. In particular, when there is a pattern of exponential decay in the autocorrelations plot, this suggests the presence of an autoregressive term, typically indicated as AR. It also suggests the series is stationary, as opposed to an autocorrelations plot that would decline linearly for example. If a comparable pattern is also visible in the partial autocorrelations plot, this would be indicative of not only the presence of an autoregressive term but also a moving average term, MA.

As can be seen on the left-hand pane of this figure, the autocorrelations collapse to insignificance. The first autocorrelation stands out, suggesting an MA(1) term. The first two partial autocorrelations lie outside the 95% confidence band, suggesting a first or second order AR process.

Figure 6-9: Autocorrelation (left-hand side) and partial autocorrelation (right-hand side) of seasonally and locally differenced pre-intervention time series



In a next step, different models of the seasonally and locally differenced time series with different ARIMA structures were tested. Three dummy variables were included to control for the outliers in August 1999, February 2000 and March 2005 (dummy_outl1, dummy_outl2 and dummy_outl3). The model with the lowest AIC and BIC values and significant AR and MA terms was chosen as the final model. Figure 6-10 contains the results of this model. This final model consists of one MA term (MA12) and two AR terms (AR1 and AR2). The AIC value of this model is -122.59 and the BIC value is -103.38.

Figure 6-10: ARIMA regression results for the pre-intervention time series

```
ARIMA regression
Sample:  1999m2 - 2008m8                         Number of obs      =       115
                                                 Wald chi2(6)       =    113.42
Log pseudolikelihood =  68.29659                 Prob > chi2        =    0.0000
------------------------------------------------------------------------------
DS12.        |              Semirobust
lnCharge_alc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lnCharge_alc |
 dummy_outl1 |    .032556   .0835937     0.39   0.697    -.1312847    .1963966
 dummy_outl2 |  -.1020674   .1104666    -0.92   0.356    -.3185781    .1144432
 dummy_outl3 |  -.1233128   .0538023    -2.29   0.022    -.2287634   -.0178621
ARMA         |
          ar |
         L1. |  -.6789599   .1107106    -6.13   0.000    -.8959487    -.461971
         L2. |  -.3123224   .1148969    -2.72   0.007    -.5375161   -.0871286
             |
          ma |
        L12. |  -.9243891   .4526919    -2.04   0.041    -1.811649   -.0371294
-------------+----------------------------------------------------------------
      /sigma |   .1218892   .0232062     5.25   0.000     .0764059    .1673725
```

A series of diagnostic test were conducted to study model fit of this final model. First, Figure 6-11 shows that the final model fits the data reasonably well. Second, Figure 6-12 contains the results of a white noise test of the residuals of this final model. This test confirmed the residuals are indeed distributed according to a white noise pattern. Bartlett's statistic did not reject the null hypothesis of white noise (0.74; p=0.64) and all the dots in the figure are within the confidence bounds. According to Figure 6-13 and Figure 6-14, the residuals are also normally distributed, another indication that the assumptions of the model are satisfied. Although the Portmanteau-Q test (66.13, p= 0.0058) rejected the hypothesis of white noise, other tests like Shaphiro-Wilk-W (-0.03, p= 0.512) and Skewness/Kurtosis (03.24, p=0.198) confirmed the normality assumption (see Figure 6-15).

Figure 6-11: Log-transformed time series and one-step-ahead predictions according to the ARIMA estimated model for the pre-intervention data

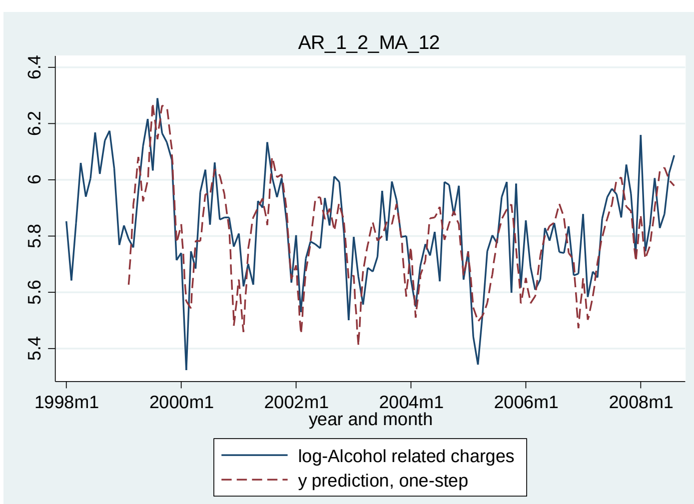


Figure 6-12: Cumulative periodogram white noise test and Bartlett's statistic of the residuals of the pre-intervention model

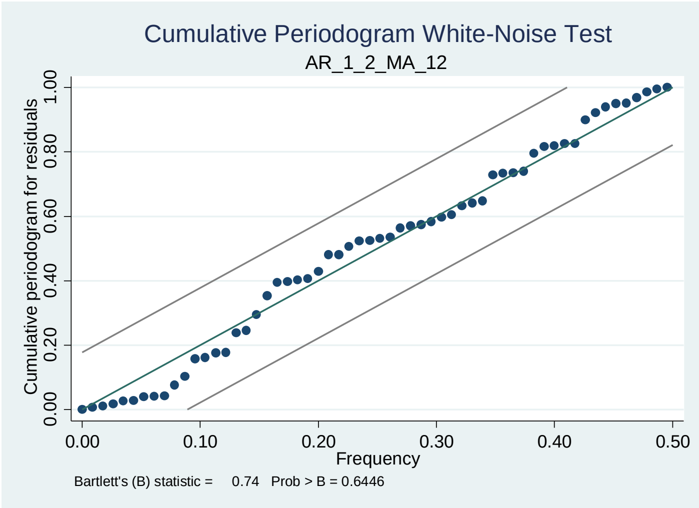


Figure 6-13: Standardized normal probability plot of the residuals of the pre-intervention model

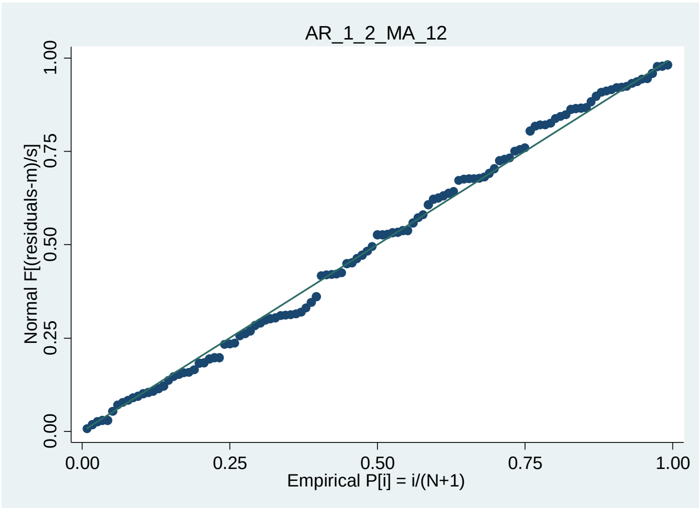


Figure 6-14: Normal density plot of the residuals of the final pre-intervention model

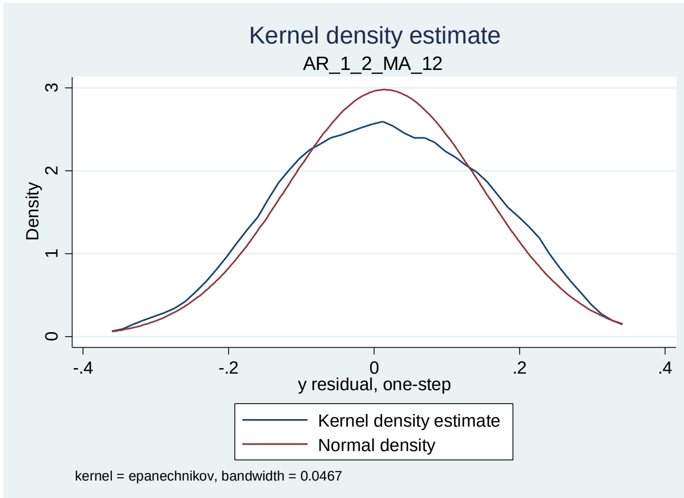


Figure 6-15: Portmanteau, Shaphiro-Wilk and Skewness/Kurtosis tests for white noise

```
Portmanteau test for white noise
Portmanteau (Q) statistic =     66.1337
 Prob > chi2(40)           =      0.0058
                   Shapiro-Wilk W test for normal data
    Variable |     Obs       W          V         z       Prob>z
-------------+--------------------------------------------------
   residuals |     115    0.98937      0.987    -0.030    0.51193
                    Skewness/Kurtosis tests for Normality
                                                       ------- joint ------
    Variable |     Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)    Prob>chi2
-------------+-----------------------------------------------------------------
   residuals |     115      0.8258         0.0776         3.24          0.1982
```

Final Model

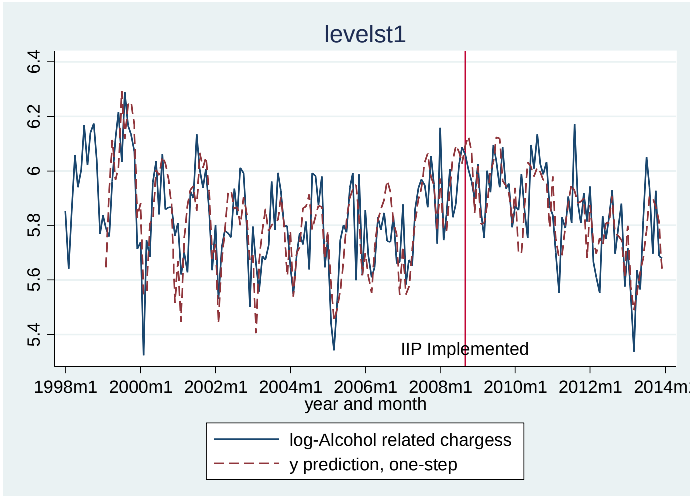
Using the described methodology, 39 models were tested. They were compared using the Akaike and Bayesian information criteria statistics (AIC and BIC). The best fitting model was the sudden temporary model with effects starting in the first month after the implementation (see Figure 6-16). Both AIC and BIC values were the lowest of all the models (AIC=-216.16; BIC=-174.73).

Figure 6-16: Sudden temporary model with an effect in the 1st month following the intervention

```
ARIMA regression
 Sample:  1999m2 - 2013m12                        Number of obs      =       179
                                                  Wald chi2(12)      =  1.40e+15
 Log pseudolikelihood =  121.0817                 Prob > chi2        =    0.0000
------------------------------------------------------------------------------
DS12.         |              Semirobust
 lnCharge_alc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
lnCharge_alc  |
  dummy_outl1 |  -.0531604   .0883802    -0.60   0.548    -.2263825    .1200617
  dummy_outl2 |  -.1786106   .1100888    -1.62   0.105    -.3943808    .0371595
  dummy_outl3 |  -.2140877   .0472608    -4.53   0.000     -.306717   -.1214583
        pulse |  -.1426334   .0445841    -3.20   0.001    -.2300167   -.0552501
   pop16_rate |   .0007056   .0003515     2.01   0.045     .0000166    .0013945
   unemp_rate |  -.0117254   .0059327    -1.98   0.048    -.0233533   -.0000976
  hdrink_rate |  -.0003556   .0007205    -0.49   0.622    -.0017677    .0010565
alc_sale_rate |  -.0348028   .0271665    -1.28   0.200    -.0880481    .0184426
 lnCharge_nalc |  -.0141995   .0094535    -1.50   0.133     -.032728    .0043289
--------------+----------------------------------------------------------------
ARMA          |
           ar |
          L1. |  -.7362725   .0665913   -11.06   0.000    -.8667891   -.6057559
          L2. |  -.3775676   .0606253    -6.23   0.000     -.496391   -.2587443
              |
           ma |
         L12. |         -1   2.55e-06 -3.9e+05   0.000    -1.000005    -.999995
--------------+----------------------------------------------------------------
       /sigma |   .1119255   .0056654    19.76   0.000     .1008215    .1230295
```

For this model the variable pulse is significant (coefficient=-0.143; p=0.001). This pulse variable corresponds to a sudden, albeit temporary effect associated with the first month after AIIP was implemented, i.e., September 2008. The effect suggests that, when controlling for trends in the population over age 16, unemployment rates, heavy drinking rates, alcohol sales and the non-alcohol related charges, the implementation of the program in September 2008 had a non-lasting significant effect in the first month of the implementation. This effect is a 13.32% ((exp(-0.143)-1)*100%) decrease in the number of alcohol-related charges. Figure 6-17 shows that the final model fits the data reasonably well.

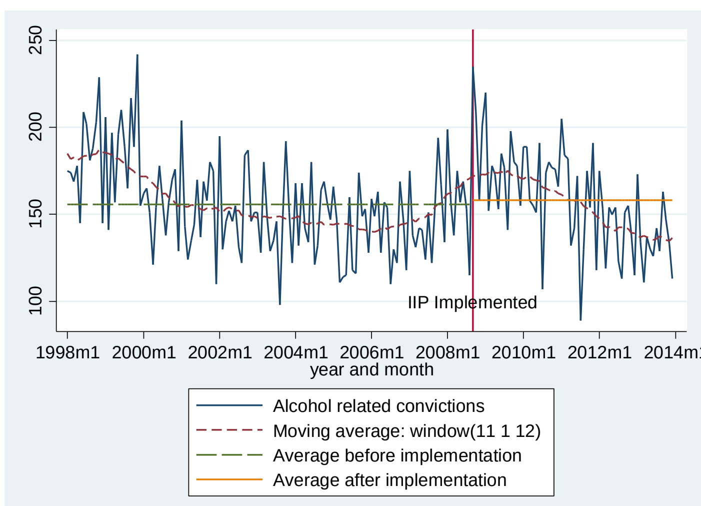
Figure 6-17: Log-transformed time series and one-step-ahead predictions according to the estimated model



6.2 Convictions

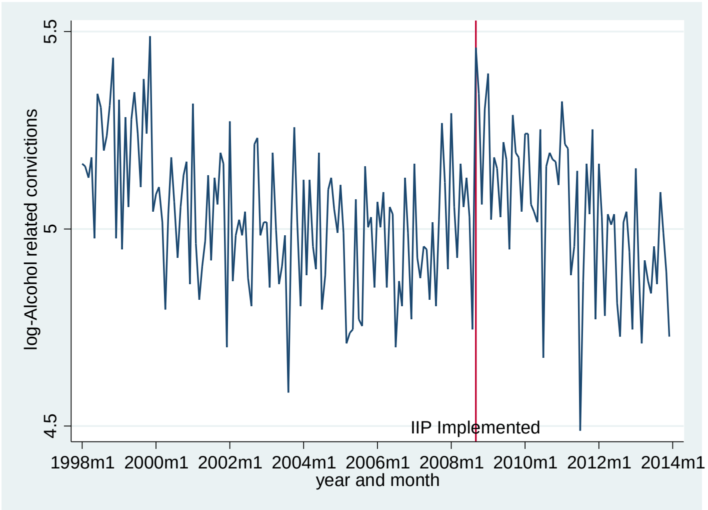
Figure 6-18 shows the time series of monthly counts of alcohol related convictions. The vertical line in this figure represents the month when AIIP was implemented (September 2008).

Figure 6-18: Monthly counts of alcohol related convictions



To mitigate the impact of outliers and changing variance over time, these monthly counts have been log-transformed. The result of this log-transformed time series is shown below.

Figure 6-19: Log-transformed series of monthly counts of alcohol related convictions



In Figure 6-20 the time series for non-alcohol related convictions used as a control series is presented and in Figure 6-21 the log-transformed series of both alcohol and non-alcohol related convictions are compared.

Figure 6-20: Monthly counts of non-alcohol related convictions

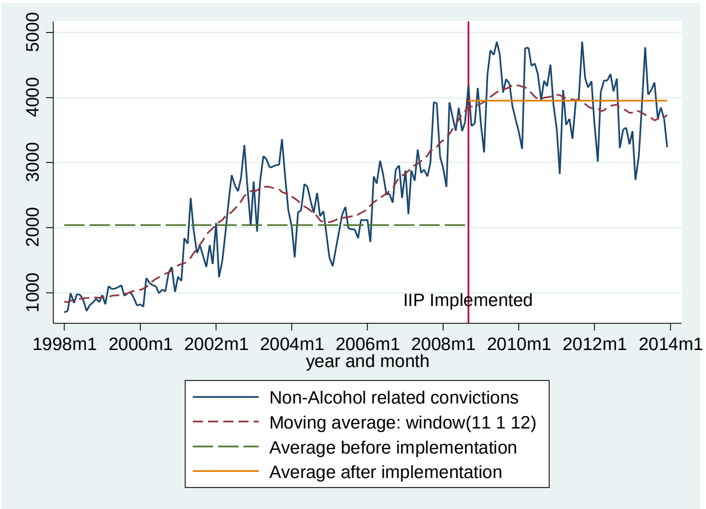
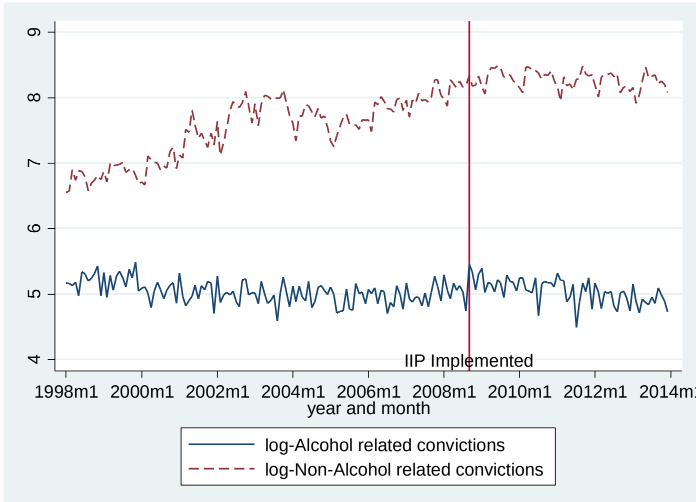


Figure 6-21: Log-transformed series of monthly counts of alcohol and non-alcohol related convictions



Model selection for pre-intervention data

When looking at the pre-intervention series (from January 1998 to August 2008), there does not appear to be an increasing trend but there does appear to be seasonal variation in the time series, suggesting seasonal differencing may be required to make this series stationary.

A regression model was used to formally test the presence of a trend in the pre-intervention log-transformed series. Figure 6-22 below contains the results from this model. As can be seen, the trend is significant, (coef.: -0.0019; s.e.: 0. 00038; p <0.001), suggesting that local differencing might be necessary.

Figure 6-22: Regression model of a trend for the pre-intervention series

```
      Source |       SS       df       MS              Number of obs =     128
-------------+------------------------------           F(  1,    126) =   23.84
       Model |  .61277406     1  .61277406             Prob > F      =  0.0000
    Residual | 3.23804855   126  .025698798            R-squared     =  0.1591
-------------+------------------------------           Adj R-squared =  0.1525
       Total | 3.85082261   127  .030321438            Root MSE      =  .16031

------------------------------------------------------------------------------
   lnConv_alc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       trend |  -.0018726   .0003835    -4.88   0.000    -.0026315   -.0011137
       _cons |   5.154671   .0285057   180.83   0.000     5.098259    5.211083
```

The time series of log-transformed counts were also standardized by creating Z-scores to identify outliers. An observation with a Z-score greater than 2.5 was considered an outlier. Two such outliers were identified for November 1999 and August 2003 (Z-scores: 2.61, 2.58). These outliers are tagged in the estimated models using dummy variables.

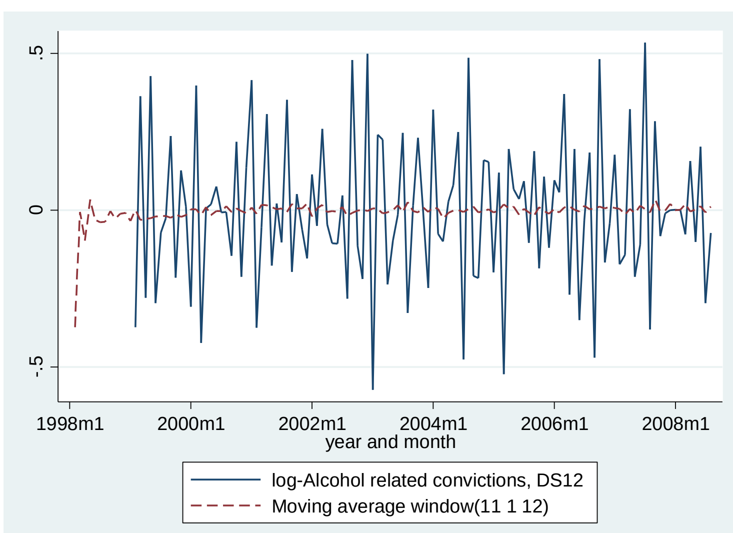
A correlogram was used to further investigate the need for seasonal differencing. The results are displayed in Figure 6-23. This correlogram confirms that there is seasonal variation in the data, suggesting seasonal differencing is required to make the time series stationary.

Figure 6-23: Correlogram of the time series

```
                                                  -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.1736   0.1776   3.9488  0.0469          |-                 |-
2        0.2123   0.1924   9.8993  0.0071          |-                 |-
3        0.3041   0.2656   22.207  0.0001          |--                |--
4        0.2558   0.1805   30.987  0.0000          |--                |-
5        0.1892   0.0695   35.831  0.0000          |-                 |
6        0.0986  -0.0683   37.158  0.0000          |                  |
7        0.1544   0.0112   40.436  0.0000          |-                 |
8        0.2704   0.2086   50.573  0.0000          |--                |-
9        0.1545   0.0993    53.91  0.0000          |-                 |
10       0.0219  -0.1102   53.978  0.0000          |                  |
11       0.1832   0.0551   58.752  0.0000          |-                 |
12       0.4683   0.4841   90.211  0.0000          |---               |---
13       0.0966   0.0371   91.561  0.0000          |                  |
14       0.1373  -0.0564   94.312  0.0000          |-                 |
15       0.1646  -0.1187     98.3  0.0000          |-                 |
16       0.1981   0.0392   104.13  0.0000          |-                 |
17       0.0948   0.0269   105.48  0.0000          |                  |
18      -0.0452  -0.0957   105.79  0.0000          |                  |
19       0.0415  -0.1447   106.05  0.0000          |                 -|
20       0.2120   0.1219   112.98  0.0000          |-                 |
21       0.0114   0.0966      113  0.0000          |                  |
22      -0.0380   0.0493   113.23  0.0000          |                  |
23       0.1314   0.0419   115.96  0.0000          |-                 |
24       0.2674   0.2071    127.4  0.0000          |--                |-
25      -0.0068  -0.0285   127.41  0.0000          |                  |
26       0.0884   0.1589   128.68  0.0000          |                  |-
27       0.0397  -0.0636   128.94  0.0000          |                  |
28       0.1228  -0.0025   131.45  0.0000          |                  |
29       0.0237  -0.0183   131.54  0.0000          |                  |
30      -0.1965  -0.2312    138.1  0.0000         -|                 -|
31       0.0156  -0.0435   138.14  0.0000          |                  |
32       0.0060  -0.1801   138.15  0.0000          |                 -|
```

The figure below shows the local and seasonal differenced log-transformed time series. The plot shows that the series seems to be stationary around a constant mean.

Figure 6-24: Local and seasonal differenced log-transformed time series and its moving average



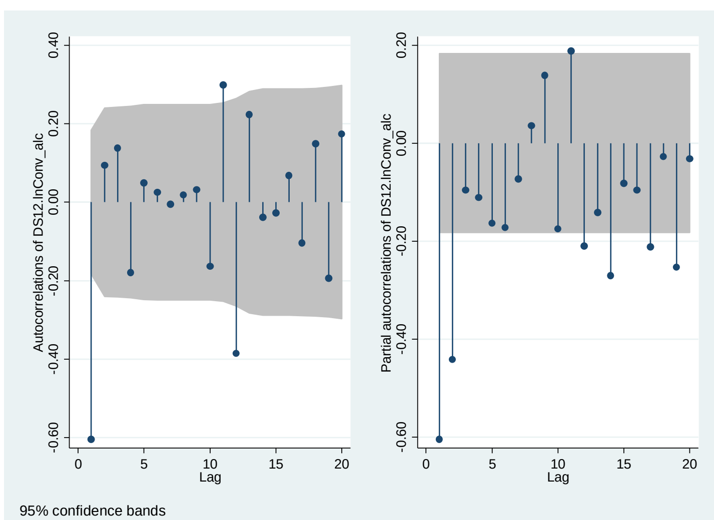
Two unit root tests (Phillips-Perron and Dickey-Fuller) rejected the null hypothesis of a unit root, thus confirming the stationarity of the local and seasonal differenced log-transformed pre-intervention time series.

Figure 6-25: Two unit root tests: Phillips-Perron and Dickey-Fuller

```
Phillips-Perron test for unit root                Number of obs   =       114
                                                  Newey-West lags =         4
                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(rho)         -161.558          -19.847           -13.728           -11.019
 Z(t)            -30.671           -3.505            -2.889            -2.579
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
Dickey-Fuller test for unit root                  Number of obs   =       114
                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(t)            -21.678           -3.505            -2.889            -2.579
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

In Figure 6-26, the autocorrelations and partial correlations of the seasonally and locally differenced log-transformed pre-intervention time series are displayed. As can be seen on the left-hand pane of this figure, the autocorrelations collapse to insignificance. The first autocorrelation stands out, suggesting an MA(1) term. The first two partial autocorrelations lie outside the 95% confidence band, suggesting a first or second order AR process.

Figure 6-26: Autocorrelation (left-hand side) and partial autocorrelation (right-hand side) of seasonally and locally differenced pre-intervention time series



In a next step, different models of the seasonally and locally differenced time series with different ARIMA structures were tested. Two dummy variables were included to control for the outliers in November 1999 and August 2003 (dummy_outl1 and dummy_outl2). The model with the lowest

AIC and BIC values and significant AR and MA terms was chosen as the final model. Figure 6-27 contains the results of this model. This final model consists of one MA term (MA12) and two AR terms (AR1 and AR2). The AIC value of this model is -109.58 and the BIC value is -93.11.

Figure 6-27: ARIMA regression results for the pre-intervention time series

```
ARIMA regression

Sample:  1999m2 - 2008m8                        Number of obs      =       115
                                                Wald chi2(5)       =    209.05
Log pseudolikelihood =  60.78874                Prob > chi2        =    0.0000
------------------------------------------------------------------------------
DS12.        |              Semirobust
  lnConv_alc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lnConv_alc   |
 dummy_outl1 |  -.0129275   .0655899    -0.20   0.844    -.1414813    .1156263
 dummy_outl2 |  -.0033706   .0673264    -0.05   0.960    -.1353279    .1285867
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.8357238    .078937   -10.59   0.000    -.9904375   -.6810101
         L2. |  -.4656441   .0848696    -5.49   0.000    -.6319854   -.2993027
             |
          ma |
        L12. |  -.7846339   .0934478    -8.40   0.000    -.9677882   -.6014795
-------------+----------------------------------------------------------------
      /sigma |   .1352065   .0087645    15.43   0.000     .1180283    .1523847
```

A series of diagnostic test were conducted to study model fit of this final model. First, Figure 6-28 shows that the final model fits the data reasonably well. Second, Figure 6-29 contains the results of a white noise test of the residuals of this final model. This test confirmed the residuals are indeed distributed according to a white noise pattern. Bartlett's statistic did not reject the null hypothesis of white noise (0.72; p=0.68) and all the dots in Figure 6-29 are within the confidence bounds. According to Figure 6-30 the residuals are also normally distributed, but in Figure 6-31 the curve does not seem to be a very clear normal distribution. Although the Portmanteau-Q test (56.74, p= 0.041) rejected the hypothesis of white noise, other tests, like Shaphiro-Wilk-W (1.33, p= 0.09) and Skewness/Kurtosis (1.86, p=0.394) did not reject the normality assumption (see Figure 6-32).

Figure 6-28: Log-transformed time series and one-step-ahead predictions according to the ARIMA estimated model for the pre-intervention data

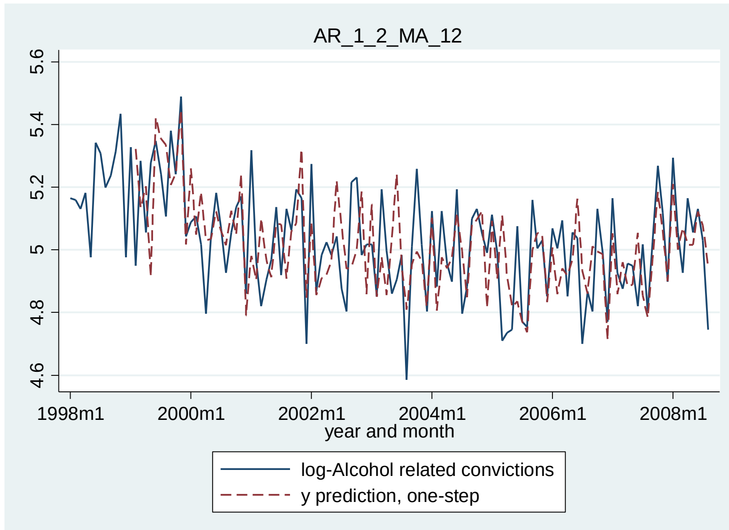


Figure 6-29: Cumulative periodogram white noise test and Bartlett's statistic of the residuals of the pre-intervention model

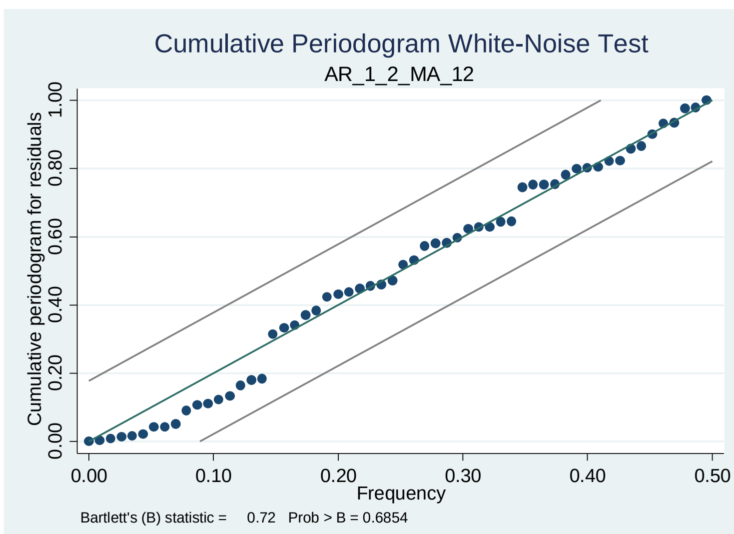


Figure 6-30: Standardized normal probability plot of the residuals of the pre-intervention model

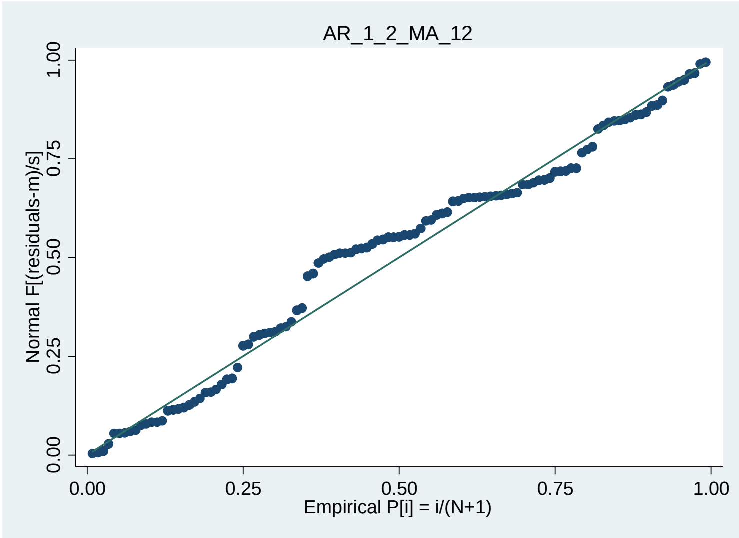


Figure 6-31: Normal density plot of the residuals of the final pre-intervention model

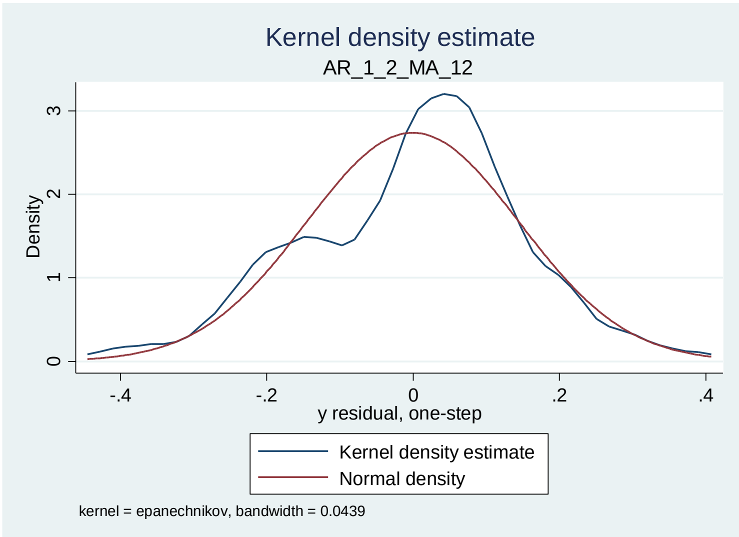


Figure 6-32: Portmanteau, Shaphiro-Wilk and Skewness/Kurtosis tests for white noise

```
Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    56.7439
 Prob > chi2(40)           =     0.0416

                   Shapiro-Wilk W test for normal data
    Variable |    Obs       W           V         z       Prob>z
-------------+--------------------------------------------------
   residuals |    115    0.98046      1.814     1.332    0.09150

                    Skewness/Kurtosis tests for Normality
                                                      ------- joint ------
    Variable |    Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)    Prob>chi2
-------------+---------------------------------------------------------------
   residuals |    115      0.1970         0.6900         1.86         0.3941
```

Final model

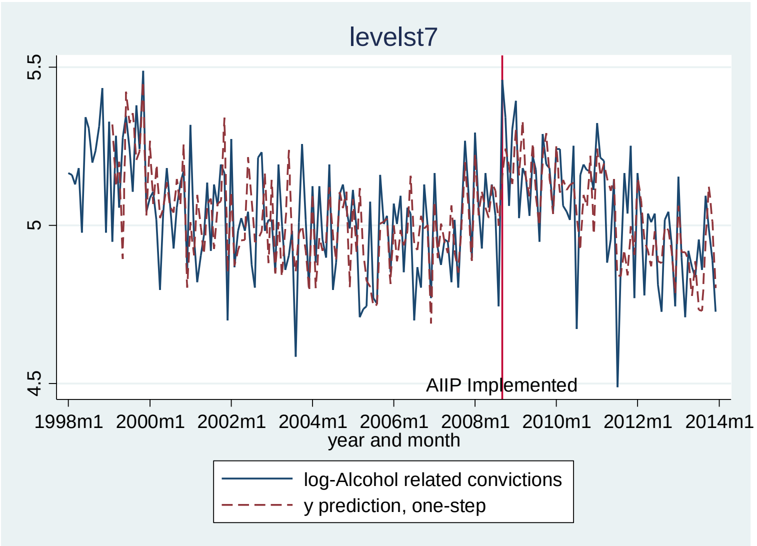
The best fitting model was the sudden temporary model with effects starting in the seventh month after the implementation (see Figure 6-33). Both AIC and BIC values were the lowest of all the models (AIC= -176.44; BIC= -138.19). For this model the variable pulse is significant (coefficient= -0.10; p=0.035). This pulse variable corresponds to a sudden, albeit temporary effect associated with the seventh month after AIIP was implemented, i.e., March 2009. The effect suggests that, when controlling for trends in the population over age 16, unemployment rates, heavy drinking rates, alcohol sales and the non-alcohol related convictions, the implementation of the program in September 2008 had a non-lasting significant effect in the seventh month of the implementation. This effect is a 9.93% decrease ((exp(-0.1045561)-1)*100) in the number of alcohol-related convictions.

It is interesting to note that while the effect found for the time series of charges was after the first month following the intervention, for the convictions the effect was found after the seventh month. It takes longer for convictions to appear compared to charges, thus it can be argued that it is normal to only see an effect several months later, in this case seven.

Figure 6-33: Sudden temporary model with an effect in the 7th month following the intervention

```
Model levelst7
RIMA regression
Sample:  1999m2 - 2013m12                        Number of obs      =       179
                                                 Wald chi2(11)      =    405.57
Log pseudolikelihood =   100.222                 Prob > chi2        =    0.0000
-------------------------------------------------------------------------------
DS12.         |              Semirobust
   lnConv_alc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
lnConv_alc    |
  dummy_outl1 |  -.0015435   .0747819    -0.02   0.984    -.1481134    .1450264
  dummy_outl2 |   .0585632   .0726081     0.81   0.420    -.0837461    .2008725
        pulse |  -.1045561   .0496276    -2.11   0.035    -.2018245   -.0072878
   pop16_rate |  -.0003131   .0003824    -0.82   0.413    -.0010626    .0004363
   unemp_rate |   .0027116    .006655     0.41   0.684     -.010332    .0157552
  hdrink_rate |   .0009446   .0006763     1.40   0.163    -.0003809    .0022701
alc_sale_rate |   .0312155   .0301193     1.04   0.300    -.0278174    .0902483
   lnConv_nalc |  -.0224424   .0101318    -2.22   0.027    -.0423004   -.0025845
--------------+----------------------------------------------------------------
ARMA          |
           ar |
          L1. |  -.8708621   .0640319   -13.60   0.000    -.9963624   -.7453618
          L2. |  -.5345373   .0608573    -8.78   0.000    -.6538154   -.4152593
              |
           ma |
         L12. |  -.8733055    .094704    -9.22   0.000    -1.058922   -.6876892
--------------+----------------------------------------------------------------
       /sigma |   .1313988   .0078315    16.78   0.000     .1160494    .1467483
```

Figure 6-34: Log-transformed time series and one-step-ahead predictions according to the estimated model



6.3 Crashes

Figure 6-35 shows the time series of monthly counts of alcohol related crashes. The vertical line in this figure represent the month when AIIP was implemented (September 2008). In Figure 6-36 the time series of both alcohol and non-alcohol related crashes are compared.

Figure 6-35: Monthly counts of alcohol related crashes

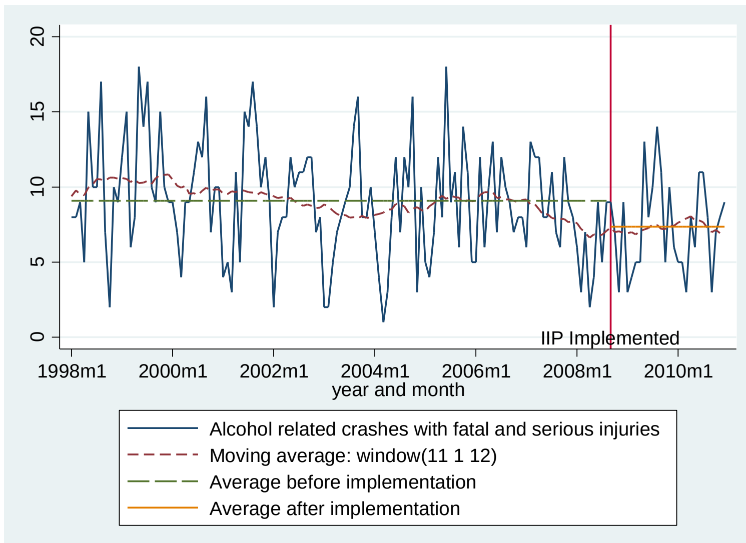
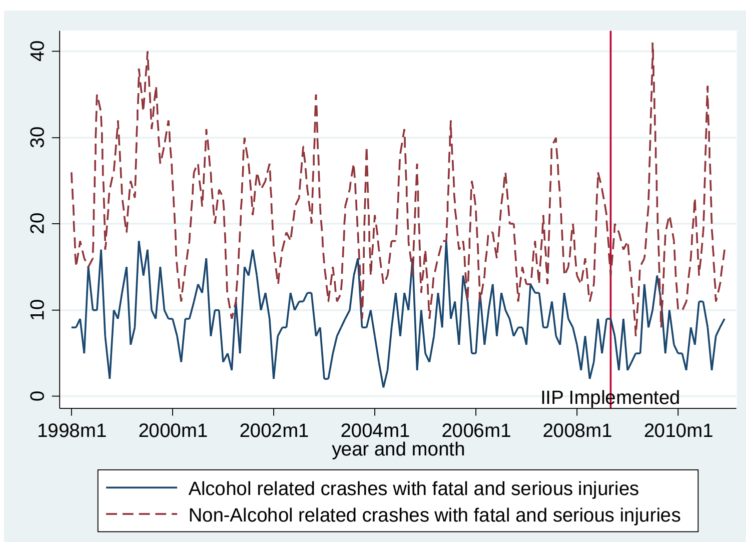


Figure 6-36: Monthly counts of alcohol and non-alcohol related crashes



Model selection for pre-intervention data

When looking at the pre-intervention series (from January 1998 to August 2008), there does not appear to be an increasing trend but there does appear to be seasonal variation in the time series, suggesting seasonal differencing may be required to make this series stationary.

A regression model was used to formally test the presence of a trend in the pre-intervention series. Figure 6-37 below contains the results from this model. As can be seen, the trend is significant, (coef.: -0.02; s.e.: 0.009; p =0.027), suggesting that local differencing might be necessary.

Figure 6-37: Regression model of a trend for the pre-intervention series

```
      Source |       SS       df       MS              Number of obs =     128
-------------+------------------------------           F(  1,   126) =    4.99
       Model |  69.041855     1   69.041855           Prob > F      =  0.0273
    Residual | 1745.01283   126  13.8493082           R-squared     =  0.0381
-------------+------------------------------           Adj R-squared =  0.0304
       Total | 1814.05469   127  14.2838952           Root MSE      =  3.7215
------------------------------------------------------------------------------
CrashInj_alc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       trend |  -.0198767   .0089023    -2.23   0.027    -.0374942   -.0022593
       _cons |   10.36799   .6617421    15.67   0.000     9.058419    11.67756
```

The time series of counts were also standardized by creating Z-scores to identify outliers. An observation with a Z-score greater than 2.5 was considered an outlier. No such Z-scores were identified.

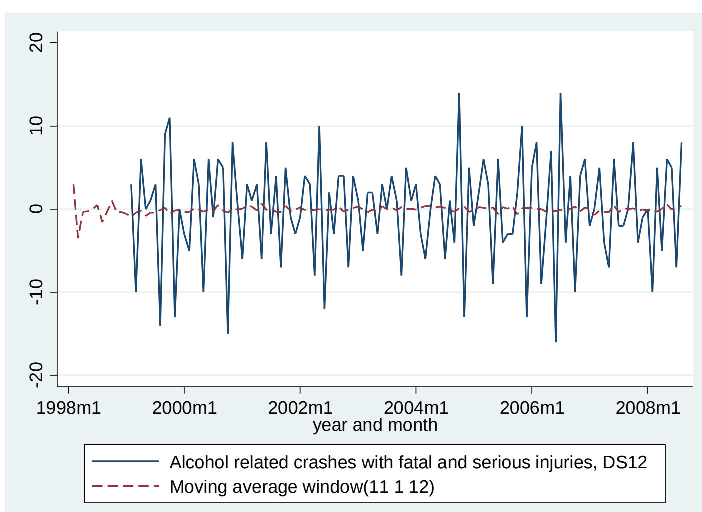
A correlogram was used to further investigate the need for seasonal differencing. The results are displayed in Figure 6-38. This correlogram confirms that there is considerable seasonal variation in the data, suggesting seasonal differencing is required to make the time series stationary.

Figure 6-38: Correlogram of the time series

```
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.2723   0.2723   9.7148  0.0018          |--                |--
2        0.2026   0.1403   15.136  0.0005          |-                 |-
3        0.0819  -0.0041   16.028  0.0011          |                  |
4       -0.1324  -0.1993   18.382  0.0010         -|                 -|
5       -0.2135  -0.1795    24.55  0.0002         -|                 -|
6       -0.2185  -0.1016   31.065  0.0000         -|                  |
7       -0.3252  -0.2306   45.606  0.0000        --|                 -|
8       -0.0885   0.0680   46.691  0.0000          |                  |
9        0.0090   0.0734   46.702  0.0000          |                  |
10       0.1762   0.1708   51.081  0.0000          |-                 |-
11       0.2916   0.1779   63.175  0.0000          |--                |-
12       0.2843   0.0871   74.769  0.0000          |--                |
13       0.2314   0.0395   82.518  0.0000          |-                 |
14       0.1728   0.0253   86.878  0.0000          |-                 |
15      -0.0754  -0.1354   87.716  0.0000          |                 -|
16      -0.0959   0.0194   89.083  0.0000          |                  |
17      -0.1940   0.0384   94.728  0.0000         -|                  |
18      -0.2405  -0.0113   103.48  0.0000         -|                  |
19      -0.2736  -0.1574   114.91  0.0000        --|                 -|
20      -0.1177  -0.0365   117.04  0.0000          |                  |
21      -0.0020   0.0213   117.04  0.0000          |                  |
22       0.1361   0.0279   119.95  0.0000          |-                 |
23       0.1889   0.0216   125.61  0.0000          |-                 |
24       0.1878  -0.0756   131.25  0.0000          |-                 |
25       0.2810   0.1905   144.01  0.0000          |--                |-
26       0.1278   0.0358   146.68  0.0000          |-                 |
27      -0.0169  -0.0635   146.72  0.0000          |                  |
28      -0.0449   0.1189   147.06  0.0000          |                  |
29      -0.1755   0.0454   152.23  0.0000         -|                  |
30      -0.1799   0.0104   157.73  0.0000         -|                  |
31      -0.1840  -0.0497   163.54  0.0000         -|                  |
32      -0.1824  -0.1337    169.3  0.0000         -|                 -|
```

The figure below (Figure 6-39) shows the local and seasonal differenced time series. The plot shows that the series seems to be stationary around a constant mean.

Figure 6-39: Local and seasonal differenced time series and its moving average



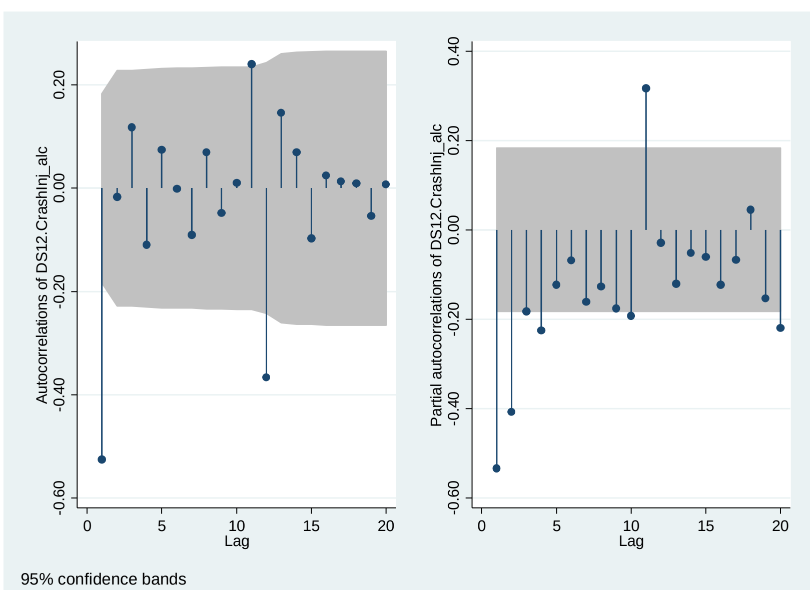
Two unit root tests (Phillips-Perron and Dickey-Fuller) reject the null hypothesis of a unit root, thus confirm the stationarity of the local and seasonal differenced series.

Figure 6-40: Two unit root tests: Phillips-Perron and Dickey-Fuller

```
Phillips-Perron test for unit root                   Number of obs    =       114
                                                     Newey-West lags =          4

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic          Value             Value             Value
------------------------------------------------------------------------------
 Z(rho)         -147.038          -19.847           -13.728           -11.019
 Z(t)            -27.774           -3.505            -2.889            -2.579
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000


Dickey-Fuller test for unit root                     Number of obs    =       114

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic          Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -19.021           -3.505            -2.889            -2.579
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

In Figure 6-41, the autocorrelations and partial correlations are displayed. As can be seen on the left-hand pane of this figure, the autocorrelations collapse to insignificance. The first autocorrelation stands out, suggesting an MA(1) term. The first two partial autocorrelations lie outside the 95% confidence band, suggesting a first or second order AR process.

Figure 6-41: Autocorrelation (left-hand side) and partial autocorrelation (right-hand side) of seasonally and locally differenced pre-intervention time series



In a next step, different models of the locally and seasonally differenced time series with different ARMA structures were tested. The model with the lowest AIC and BIC values and significant AR and MA terms was chosen as the final model. Figure 6-42 contains the results of this model. This final model consists of one MA term (MA1) and one AR term (AR12). The AIC value of this model is 661.1 and the BIC value is 669.34.

Figure 6-42: ARIMA regression results for the pre-intervention time series

```
Model AR_12_MA_1
ARIMA regression

Sample:  1999m2 - 2008m8                        Number of obs      =       115
                                                Wald chi2(2)       =  4.02e+11
Log pseudolikelihood = -327.5517                Prob > chi2        =    0.0000
------------------------------------------------------------------------------
DS12.        |              Semirobust
CrashInj_alc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ar |
        L12. |  -.4267026   .0916262    -4.66   0.000    -.6062867   -.2471185
             |
          ma |
         L1. |  -1.000001   1.62e-06 -6.2e+05   0.000    -1.000004   -.9999976
-------------+----------------------------------------------------------------
      /sigma |   4.036472   .2522877    16.00   0.000     3.541997    4.530947
```

A series of diagnostic test were conducted to study model fit of this final model. First, Figure 6-43 shows that the final model fits the data reasonably well. Second Figure 6-44 contains the results of a white noise test of the residuals of this final model. This test confirms the residuals are indeed distributed according to a white noise pattern. Bartlett's statistic did not reject the Null hypothesis of white noise (0.52; p=0.95) and all the dots are within the confidence bounds.

Figure 6-43: Time series and one-step-ahead predictions according to the ARIMA estimated model for the pre-intervention data

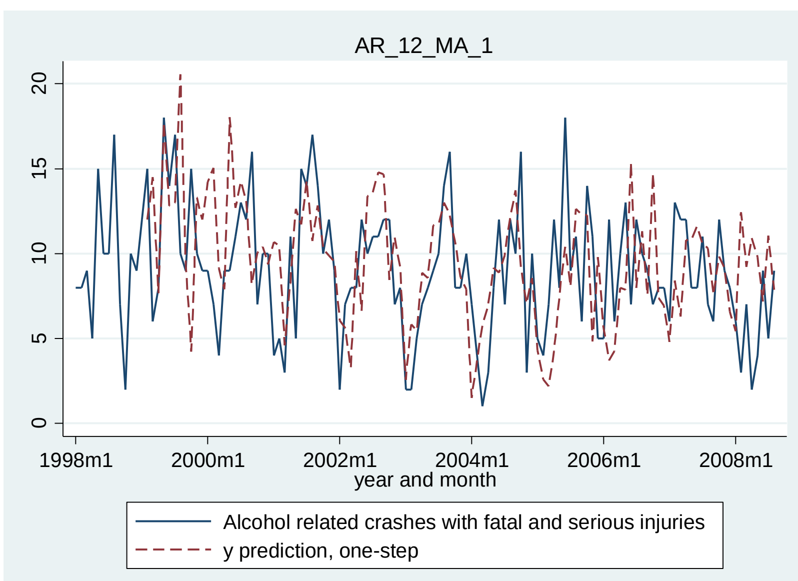
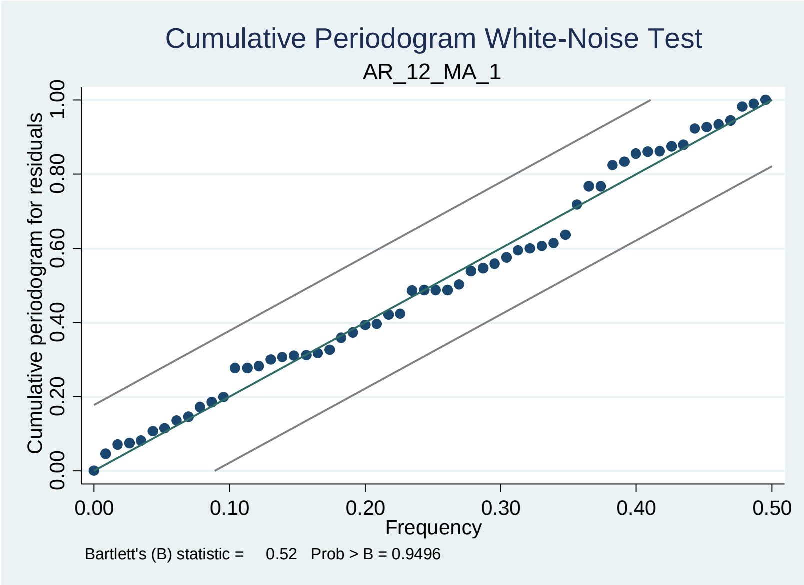


Figure 6-44: Cumulative periodogram white noise test and Bartlett's statistic of the residuals of the pre-intervention model



According to Figure 6-45 and Figure 6-46, the residuals are also normally distributed, another indication that the assumptions of the model are satisfied. Other tests Portmanteau-Q (29.1, p=0.899), Shaphiro-Wilk-W (-0.526, p=0.7) and Skewness/Kurtosis (0.04, p=0.98) also confirm the normality assumption (see Figure 6-47).

Figure 6-45: Standardized normal probability plot of the residuals of the pre-intervention model

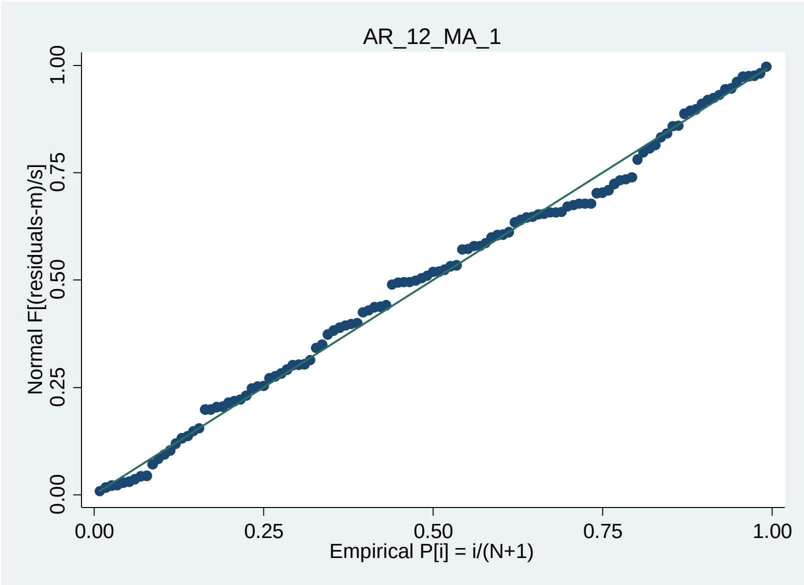


Figure 6-46: Normal density plot of the residuals of the final pre-intervention model

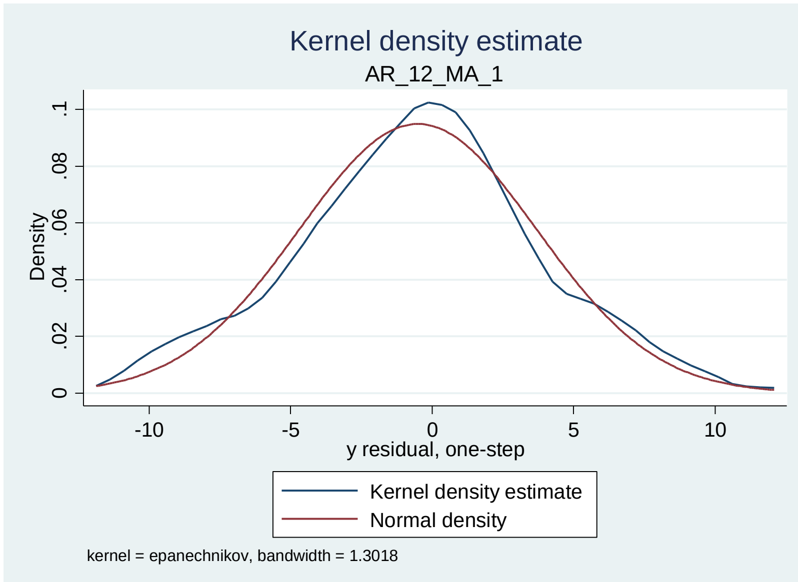


Figure 6-47: Portmanteau, Shaphiro-Wilk and Skewness/Kurtosis tests for white noise

```
Portmanteau test for white noise
Portmanteau (Q) statistic =    29.1055
 Prob > chi2(40)          =     0.8987

                   Shapiro-Wilk W test for normal data
    Variable |    Obs       W           V         z       Prob>z
-------------+--------------------------------------------------
   residuals |    115    0.99149      0.790    -0.526    0.70070
                    Skewness/Kurtosis tests for Normality
                                                    ------- joint ------
    Variable |    Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)    Prob>chi2
-------------+---------------------------------------------------------------
   residuals |    115      0.9485         0.8517         0.04          0.9806
```

Final Model

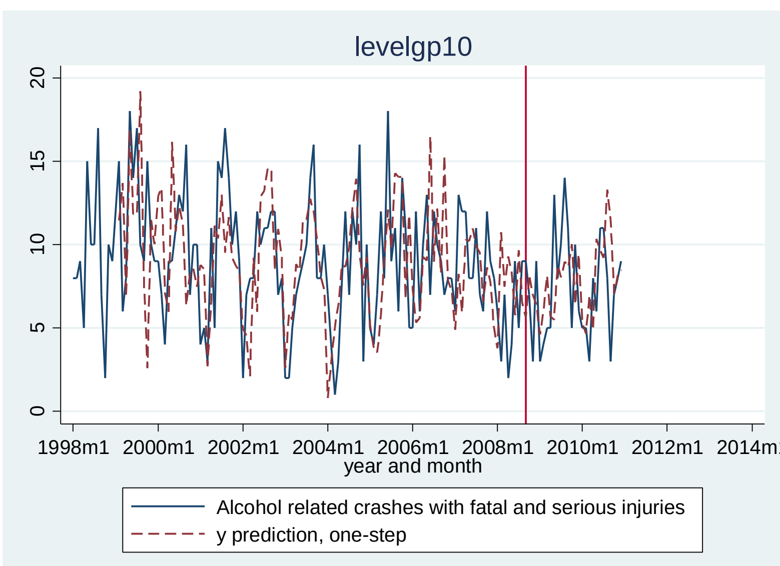
The best fitting model was the gradual permanent model with effects starting in June 2009, the tenth month after the implementation (see Figure 6-48). Both AIC and BIC values were the lowest of all the models (AIC=811.675; BIC=835.378). For this model the variable TIIPimpl is not significant at the 5% level but it is significant at the 10% level (coefficient=-0.0025; p=0.083).

Figure 6-48: Gradual permanent model with an effect in the 10th month of the program

```
Model levelgp10
ARIMA regression
Sample:  1999m2 - 2010m12                        Number of obs      =       143
                                                 Wald chi2(7)       =     52.04
Log pseudolikelihood = -397.8378                 Prob > chi2        =    0.0000
------------------------------------------------------------------------------
DS12.        |             Semirobust
 CrashInj_alc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
CrashInj_alc |
     TIIPimpl |  -.0025326   .0014632    -1.73   0.083    -.0054004    .0003352
   pop16_rate |  -.0109598   .0066544    -1.65   0.100    -.0240021    .0020825
   unemp_rate |   .1147382   .0660511     1.74   0.082    -.0147195     .244196
  hdrink_rate |  -.0170397   .0104005    -1.64   0.101    -.0374243     .003345
alc_sale_rate |   1.328277   .6782373     1.96   0.050    -.0010436    2.657598
CrashInj_nalc |  -.0028009   .0203458    -0.14   0.891    -.0426779    .0370762
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L12. |  -.4291277   .0736654    -5.83   0.000    -.5735094   -.2847461
              |
          ma |
          L1. |         -1   .0000245 -4.1e+04   0.000    -1.000048   -.9999519
-------------+----------------------------------------------------------------
       /sigma |   3.799657   .2177625    17.45   0.000      3.37285    4.226464
```

The results suggest that, when controlling for trends in the population over age 16, unemployment rates, heavy drinking rates, alcohol sales and the non-alcohol related fatal and serious crashes, the implementation of the program in September 2008 did not have a significant effect at the 5% level. However, the effect at the 10% level of significance represents a decrease of 0.0025 in the number of alcohol-related crashes every month since June 2009. Note that, from a statistical point of view, this represents a small decrease corresponding to one fatal or serious alcohol-related crash in approximately 33 years (1/0.0025326/12).

Figure 6-49: Time series and one-step-ahead predictions according to the estimated model



6.4 Conclusions

The possible impact of the interlock program was examined in terms of charges, convictions and crashes using time series analysis. With respect to charges and convictions, the results presented in sections 6.1 and 6.2 suggest that there are no permanent effects associated with the implementation of the program in the number of alcohol-related charges and convictions. There are significant, albeit temporary effects in the first and seventh month of the implementation. These effects are a 13.32% decrease in the number of alcohol-related charges and a 9.93% decrease in the number of alcohol-related convictions in the first and seventh month of the implementation of the program respectively.

With respect to crashes, according to the results presented in section 6.3, there are no significant effects associated with the implementation of the program in the number of alcohol-related crashes with fatal and serious injuries at the 5% level of statistical significance. However, there is a gradual permanent effect at the 10% level of significance that represents a small decrease – statistically speaking – of 0.0025 in the number of alcohol-related crashes every month since June 2009 (10th month of the program). Note that this is perhaps not unexpected as to date, most studies have not yet been able to definitively demonstrate a positive impact on crashes due to the lack of sufficient data.

7. QUESTIONNAIRE DATA ANALYSIS

Questionnaire data analysis was used to study attitudes and opinions regarding the interlock program, drinking behaviour, and drinking driving behaviour at different moments of time. Of the 481 participants who provided consent to participate and filled out questionnaires at intake, 32 filled out exit questionnaires (exiting the DWI program and the interlock program) and only 20 filled out the six months follow-up questionnaire. Furthermore, although there were 20 questionnaires at follow-up time for interlock participants, only one participant completed questionnaires at the three different times (intake, exit and follow-up).

Consequently, description of the aggregated data at each relevant time is presented but analysis of the data to study behavioural changes is very limited. Regression models to describe the change in RIASI scores between intake and exit are presented for the 26 participants for whom data were available, but caution is warranted when interpreting these results.

7.1 Descriptive analysis

Questionnaires completed by participants assessed a variety of characteristics. A series of statistical tests were conducted to compare the observed differences between the different groups. The Pearson's chi-squared test (X^2) was used to compare the observed frequencies (e.g., frequencies of a specific answer to a question) for the groups when those values were sufficiently large (above 5), otherwise the Fisher's exact test was used instead. The t-test was used for pairwise comparison of mean values between two of the groups. Table 7-1 presents measures that assess alcohol consumption and drink driving characteristics of both interlock groups and control group participants at intake. The data analyses considered missing data when the respondents did not answer a question, except in the RIASI questionnaire. According to the RIASI manual inconsistent or missing items should be considered as a positive response (Nochajski, 2002). The p-values of the statistics tests are presented in the table. When no subscript is used the comparison was between all three groups, whereas a subscript 1 represents the comparison between both interlock groups (mandatory and voluntary). Similarly, a subscript 2 represents the comparison between both voluntary groups (interlock and control).

Table 7-1: Measures of Alcohol Consumption and Drink Driving at intake

	Interlock-M	Interlock-V	Control-V	Significance Tests
# times impaired, needed to drive but didn't (Q1 self-report)				
Mean (SD)	3 (0.63)	2.17 (0.48)	2.28 (0.45)	t_1=1.1, df_1=9, p_1=0.3 t_2=-016, df_2=18, p_2=0.88
# times impaired, needed to drive and did it (Q2 self-report)				
Mean (SD)	2 (0)	2(0)	2.75 (1)	Insufficient data
Total (RIASI)				
Potential problems (%)	(100%)	(100%)	(99.26%)	p(F) =1

Table 7-1: Measures of Alcohol Consumption and Drink Driving at intake				
	Interlock-M	**Interlock-V**	**Control-V**	**Significance Tests**
Mean (SD)	18.74 (0.59)	18.64 (0.65)	18.16 (0.28)	t_1=0.13, df_1=143, p_1=0.9 t_2=0.72, df_2=336, p_2=0.47
Risk of Recidivism (RIASI)				
Potential problems (%)	(98.7%)	(95.4%)	(98.9%)	p(F) =0.12
Mean (SD)	6.52(0.21)	6.39(0.25)	6.86(0.12)	t_1=0.39, df_1=143, p_1=0.7 t_2=-1.7, df_2=336, p_2=0.09
Readiness to Change Stage				
Precontemplation: (%)	(13.04%)	(35.21%)	(27.42%)	p(F) = 0.004 p_1(F) = 0.002 p_2(F) = 0.1
Contemplation: (%)	(13.04%)	(8.45%)	(6.45%)	
Action: (%)	(55.43%)	(50.7%)	(50.65%)	
Maintenance: (%)	(18.48%)	(5.63%)	(15.48%)	

To assess the drinking driving behaviour, participants were asked to complete the self-reported behaviour questionnaire. Over 90 percent of the participants in each group reported never being in the situation where they need to drive their car while impaired and decided not to drive it. Of those who did report being in such situation, the interlock-mandatory group has the largest mean (3 times) for the number of occasions but the differences are not significant (p-values>0.05). With respect to the situation where they need to drive their car while impaired and decided to drive it, only 2% of the respondents were in such a situation and the data were insufficient to compare groups.

The RIASI contains several subscales, two of which are of relevance here – Total score, and Risk of Recidivism. The RIASI total scores were in the wide range of 8-41. A total score of 9 or more is recommended as a cut-off point for an indication of potential problems. Of all participants, 99.8% had a total score above 9, the average total score was 18.34 and did not differ significantly between the different groups (p-values>0.05).

The Risk of Recidivism subscale assesses the likelihood that the individual will be arrested on a subsequent occasion for a drink driving offence. A score of 3 or more for males and 4 or more for females on this scale is considered indicative of a high risk of recidivism. Overall, 98.32% of participants exhibit this level of risk. The average score was 6.72 and it did not differ significantly between the different groups (p-values>0.05).

The Readiness to Change questionnaire allowed participants to be placed in one of four categories based on the Prochaska and DiClemente's model of stages of change. The stage indicates the individual's progress towards a change in their problematic behaviour. Overall, approximately one-quarter (25.79%) of participants were in the pre-contemplation stage. Essentially, these individuals were not considering any change in their behaviour. Approximately 8% were considering a change in their behaviour but had done little if anything (contemplation). The stage with more participants (51.6%) is the action stage in which they were actively engaged in changing their behaviour. Finally, 14.6% were considered to be in the maintenance stage where they had changed their behaviour and were working to maintain the changes they had made. Both interlock groups differ in the assessed stage of change (statistically significant, p-value=0.002). The largest difference is in the pre-contemplation stage, where there was a larger proportion of interlock-voluntary (35.2%) compared to interlock-mandatory participants (13%). Also the total percentage of participants in the action and maintenance stages is larger (73.91%) in the interlock-mandatory than in the interlock-voluntary (56.33%). However, the differences between both voluntary groups are not statistically significant (p-value=0.1).

Table 7-2 presents measures that assess alcohol consumption and drink driving characteristics of both interlock groups and control group participants at exit

Table 7-2: Measures of Alcohol Consumption and Drink Driving at exit

	Interlock-M	Interlock-V	Control-V	Significance Tests
# times impaired, needed to drive but didn't (Q1 self-report)				
Mean (SD)	0	0	2.5 (1.5)	Insufficient data
# times impaired, needed to drive and did it (Q2 self-report)				
Mean (SD)	0	0	0	Insufficient data
Total (RIASI)				
Potential problems N (%)	(100%)	(100%)	(100%)	
Mean (SD)	17.57 (1.73)	14.3 (0.65)	17.45 (0.89)	t_1=0.97, df_1=8, p_1=0.34 t_2=-1.2, df_2=21, p_2=0.24
Risk of Recidivism (RIASI)				
Potential problems (%)	(100%)	(66.7%)	(95%)	p(F) = 0.24
Mean (SD)	6.28 (0.47)	5.7 (2.1)	6.45 (0.37)	t_1=0.43, df_1=8, p_1=0.67 t_2=-0.66, df_2=21, p_2=0.52
Readiness to Change Stage				
Precontemplation: (%)	(0%)	(25%)	(22.2%)	p(F) = 0.372 p_1(F) = 0.3 p_2(F) = 0.3
Contemplation: (%)	(0%)	(0%)	(0%)	
Action: (%)	(57.1%)	(75%)	(38.9%)	
Maintenance: (%)	(42.9%)	(0%)	(38.9%)	

One noticeable fact is that both the average total and risk of recidivism RIASI scores decrease for the three groups from intake to exit. And the decreases are larger for the interlock-voluntary group. Another interesting fact is that the percentage of participants in the pre-contemplation and contemplation stage decrease at exit, where the large majority is in the action or maintenance stages. However, the data are too limited to draw robust conclusions.

Table 7-3 presents measures that assess alcohol consumption and drink driving characteristics of both interlock groups at follow-up. An interesting fact is that both the average total and risk of recidivism RIASI scores increase for the two groups from exit to follow-up. Also, the percentage of participants in the pre-contemplation and contemplation stage increase at follow-up. However, again, the data are too limited to draw any conclusions in this respect.

Table 7-3: Measures of Alcohol Consumption and Drink Driving at follow-up			
	Interlock-M	**Interlock-V**	**Significance Tests**
# times impaired, needed to drive but didn't (Q1 self-report)			
Mean (SD)	2.5(0.5)	1(0)	Insufficient data
# times impaired, needed to drive and did it (Q2 self-report)			
Mean (SD)	0(0)	0(0)	Insufficient data
Total (RIASI)			
Potential problems (%)	(100%)	(100%)	
Mean (SD)	18.1 (1.41)	18.54 (2.81)	t_1=-0.13, df_1=18, p_1=0.9
Risk of Recidivism (RIASI)			
Potential problems (%)	(88.9%)	(90.9%)	p_1(F) = 1
Mean (SD)	6.3 (0.73)	6.45 (1)	t_1=-0.09, df_1=18, p_1=0.93
Readiness to Change Stage			
Precontemplation: (%)	(33.3%)	(63.6%)	p_1(F) = 0.41
Contemplation: (%)	(11.1%)	(0%)	
Action: (%)	(44.4%)	(18.2%)	
Maintenance: (%)	(11.1%)	(18.2%)	

Regarding the question about driving a non-interlocked vehicle while in the program, a few members (2/16) of the mandatory-interlock group reported driving a non-interlocked vehicle (1 or 2 occasions). The respondents from the voluntary-interlock group reported zero occasions driving a non-interlocked vehicle. This is an important finding since it is typically assumed that this rarely happens and it is not common to find evidence; the evidence from this study supports the notion that this rarely happens indeed.

7.2 Regression analysis

As previously mentioned, analysis of the data to study behavioural changes is very limited. Data on RIASI at intake and exit was available for only 26 participants. Therefore caution is warranted when interpreting the following results. The figure below shows a regression model for the change in RIASI total score defined as the RIASI total score at intake minus the RIASI total score at exit.

dxRITotal = RIASI_Total_intake - RIASI_Total_exit

```
      Source |       SS       df       MS              Number of obs =      26
-------------+------------------------------           F(  6,    19) =    3.01
       Model |  323.050851     6  53.8418085           Prob > F      =  0.0305
    Residual |  339.410687    19  17.8637204           R-squared     =  0.4877
-------------+------------------------------           Adj R-squared =  0.3259
       Total |  662.461538    25  26.4984615           Root MSE      =  4.2265
--------------------------------------------------------------------------------
     dxRITotal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
         group |
vol-interlock  |   2.009158   4.107533     0.49   0.630    -6.588007    10.60632
  vol-control  |   .2232597   1.997579     0.11   0.912    -3.957721     4.40424
RITotal_intake |   .9888849   .3314542     2.98   0.008     .2951433    1.682626
RIRecid_intake |  -.8823472   .9639474    -0.92   0.371    -2.899912    1.135218
           age |   .0709267   .0711321     1.00   0.331    -.0779544     .2198078
        gender |  -2.068398   3.407909    -0.61   0.551    -9.201234    5.064438
         _cons |  -10.56961   7.645717    -1.38   0.183    -26.57228    5.433055
```

The results indicate that in these data both voluntary groups have larger changes (positive coefficients) in the RIASI total score over time than the interlock mandatory group. The change is larger for the voluntary interlock group than for the control group. However, these results are not statistically significant (p-values>0.05). A significant result (p-value=0.008) is that larger scores at intake correspond to larger changes.

The figure below shows a regression model for the change in RIASI recidivism score defined as the RIASI recidivism score at intake minus the RIASI recidivism score at exit.

```
      Source |       SS       df       MS              Number of obs =      26
-------------+------------------------------           F(  6,    19) =    2.61
       Model |  42.7904254     6  7.13173756           Prob > F      =  0.0509
    Residual |  51.8249592    19  2.72762943           R-squared     =  0.4523
-------------+------------------------------           Adj R-squared =  0.2793
       Total |  94.6153846    25  3.78461538           Root MSE      =  1.6516
--------------------------------------------------------------------------------
      dxRIRecid |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+-----------------------------------------------------------------
          group |
vol-interlock   |    .087128   1.605047     0.05   0.957    -3.272274     3.44653
  vol-control   |  -.1693425   .7805678    -0.22   0.831     -1.80309    1.464405
                |
        RITotal |  -.0638466    .129518    -0.49   0.628    -.3349309    .2072378
        RIRecid |     .78526   .3766692     2.08   0.051    -.0031176    1.573638
            age |   .0249046   .0277953     0.90   0.381    -.0332717     .083081
         gender |  -2.534205   1.331664    -1.90   0.072     -5.32141    .2530004
          _cons |   .1663637   2.987617     0.06   0.956    -6.086791    6.419518
```

The results indicate that in these data the voluntary interlock group has larger changes (positive coefficient) in the RIASI recidivism score over time than the interlock mandatory group. On the other hand, the voluntary control group has smaller changes in the RIASI recidivism score over time than the interlock mandatory group. However, these results are not statistically significant (p-values>0.05). An almost significant result (p-value=0.051) is that larger scores at intake correspond to larger changes.

7.3 Conclusions

The amount of data from the questionnaires at exit and follow-up was too limited to draw meaningful conclusions about behavioural changes. This needs to be kept in mind when considering the results from these analyses.

The data from the RIASI questionnaire revealed that at the beginning of the study there were no significant differences between the groups with respect to the extent of alcohol use and recidivism. Basically all participants had potential problems with alcohol with no significant differences in the level of the problems. A large percentage of participants (98.3%) had a high risk of recidivism.

The readiness to change questionnaire at the beginning of the study shows differences in the attitudes among mandatory versus voluntary interlock participants. The majority of the mandatory participants (73.9%) were in the action or maintenance stages where they were actively changing their behaviour or were working to prevent a relapse. The percentage of voluntary-interlock participants in these stages was 56.33%. No significant differences were found in this respect between both voluntary groups.

The self-reported behaviour questionnaire revealed no significant differences between the groups. However, an interesting reported fact is that there was some evidence showing that some interlock participants (in the mandatory group) drove a non-interlocked vehicle while in the program. This is an important finding since it is typically assumed that this rarely happens. As such, the evidence from this study supports the notion that this rarely happens indeed.

8. INTERLOCK DATA ANALYSIS

8.1 Descriptive analysis

Interlock data for 1,867 participants in the AIIP over the period November 2008 until mid-July 2014 were obtained from Alcohol Countermeasure System Corp., the sole vendor in Nova Scotia. Of those 1,867 participants, 33 had multiple installations dates and were not considered in the study since it was not possible to clearly define their participation in the program. From the remaining 1,834 participants, 510 had the interlock device installed after the end of the defined intake period for the study (November 2008 – December 2012) and one did not have any reported event. Information for the 1,323 participants for whom the interlock device was installed during the intake period was used looking at all events during the tracking period of at least 18 months (until the end of June 2014).

The information about mileage driven contains values that are very different from the majority of cases. These could be considered outliers, and should be removed when calculating averages. Using standard scores of the mileage values, five outliers were identified (see table below).

Table 8-1: Outliers in the mileage information

standard-score	Km/month	mandatory/voluntary
18.73	909,796	voluntary
6.57	322,373	mandatory
27.31	1,324,181	mandatory
10.86	529,994	mandatory
6.19	304,169	mandatory

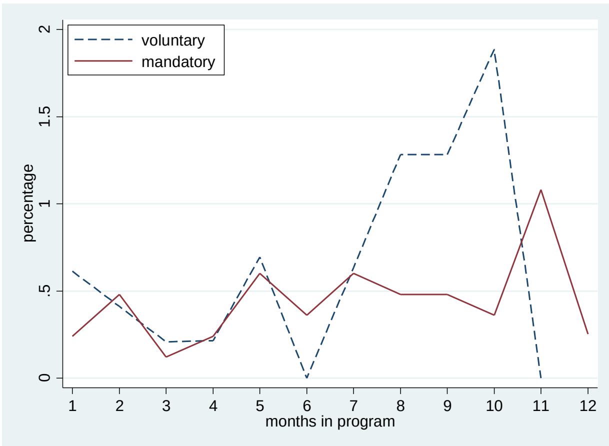
Table 8-2 compares mandatory versus voluntary participants in terms of descriptive statistics for time in the program (in months), mileage driven (in kilometers, total logged events per month and attrition rate. The attrition rate in a period represents the percentage of participants who left the program before their anticipated termination date during that period.

Table 8-2: Descriptive Characteristics of Interlock participants

	Mandatory	Voluntary
Months in program		
mean (SD)	26.1 (10.3)	10.1 (5.3)
Average* mileage driven per month		
mean (SD)	2,946 (3,893)	2,546 (4,276)
median (IQR)	1,961 (1,782)	1,690 (1,332)
Total logged events per month		
mean (SD)	268.4 (115.3)	266 (122)
median (IQR)	259.7 (147.6)	244.1 (147)
Attrition in 12 months		
	3.3%	5.3%

The data show that on average a mandatory participant drove 15.6% (100*(2946/2546-1)) more in a month than a voluntary participant. The median of mileage per month for the mandatory participants is larger (16%) than for the voluntary participants. The Wilcoxon rank-sum test rejects that the mileage per month for both groups have the same distribution (p-value<0.001). With respect to the total logged events per month, the hypothesis of equal distribution cannot be rejected (p-value=0.37). Also, with respect to the attrition rate over 12 months, the hypothesis of equal attrition rate cannot be rejected (two-sample test of proportions, p-value=0.09). The figure below shows the monthly attrition rate during the first 12 month period for each group.

Figure 8-1: Monthly attrition rate during the first 12 months in the program



A total of 7,148,274 events were collected from the logged events by the interlock devices used by the participants. The collected events are the results from the breath sample when trying to start the car (at start-up) or after having started the car (running retest). Results from these breath samples are classified according to the BAC level as "pass" (BAC level under 0.02%) or "fail" (BAC level over 0.02%). Table 8-3 shows the distribution of event types in the data.

Table 8-3: Event types

	N	%
All	7,148,274	100
fail at start-up	11,282	0.16
fail running retest	3,141	0.04
pass at start-up	3,760,015	52.60
pass running retest	3,373,836	47.20

The data have been analyzed in relation to the type of events. The events have been broken down in periods of time to help reveal changes over time. The analyses have also been broken down by gender, program status (mandatory or voluntary), condition 37 and device type. The next section presents the analyses based on periods of three months (section 8.2). This is followed by an analysis of the data by month (section 8.3).

8.2 Fails every three months period

Table 8-4, Table 8-5 and Table 8-6 show the distribution of offenders who blew over 0.02, 0.05 and 0.08 respectively. The 0.02 limit is an indication of the consumption of some amount of alcohol. The 0.05 limit is the provincial offence limit and the 0.08 is the Canadian Criminal Code limit. Each table shows the total number of participants in the program, the number of participants who blew over the specific limit with the corresponding percentages, the total number of fails over the limit and the average number of failed blows per participants at each three month-period.

The percentage of offenders who blew over each specific limit declined as they spent more time on the interlock. With respect to blows over the 0.02 limit it started at 68.9% in the first three months and ended at 45.4% in the last three months of the 30 months period (months 28-30). This represents a reduction of 34.1% (100*(68.93-45.42)/68.93) of blows over 0.02.

As for the average number of blows over 0.02 per offender, there is a steady decreasing pattern beginning at 2.43 blows per offender in the first period to 0.96 in the months 25-27. This steady decrease is followed by a small increase in the last period to 1.34 blows over 0.02 per offender.

Table 8-4: Participants who blew over 0.02

Month	Participants in the program	Participants over limit	% Participants over limit	95%-CI			# fails over limit	Average fails/per Participant
1-3	1323	912	68.93	66.36	-	71.42	3210	2.43
4-6	1298	821	63.25	60.56	-	65.88	2463	1.90
7-9	1240	667	53.79	50.97	-	56.59	1867	1.51
10-12	957	510	53.29	50.07	-	56.49	1356	1.42
13-15	819	416	50.79	47.31	-	54.27	1152	1.41
16-18	711	347	48.80	45.07	-	52.55	929	1.31
19-21	663	329	49.62	45.75	-	53.50	861	1.30
22-24	588	273	46.43	42.34	-	50.55	713	1.21
25-27	534	189	35.39	31.33	-	39.61	525	0.98
28-30	273	124	45.42	39.41	-	51.53	366	1.34

With respect to the percentage of offenders who blew over the 0.05 limit, it started at 31.97% in the first three months and ended at 15.75%. This represents a reduction of 50.8% (100*(31.97-15. 75)/31.9) of blows over 0.05. The average blows over 0.05 per offender shows a similar decreasing and increasing pattern to the blows over 0.02, in this case beginning at 0.64 blows per offender to 0.24 in the last period.

Table 8-5: Participants who blew over 0.05

Month	Participants in the program	Participants over limit	% Participants over limit	95%-CI			# fails over limit	Average fails/per Participant
1-3	1323	423	31.97	29.46	-	34.56	852	0.64
4-6	1298	333	25.65	23.30	-	28.12	566	0.44
7-9	1240	228	18.39	16.27	-	20.66	367	0.30

Table 8-5: Participants who blew over 0.05						
Month	Participants in the program	Participants over limit	% Participants over limit	95%-CI	# fails over limit	Average fails/per Participant
10-12	957	164	17.14	14.80 - 19.68	309	0.32
13-15	819	131	16.00	13.55 - 18.69	248	0.30
16-18	711	105	14.77	12.24 - 17.59	181	0.25
19-21	663	89	13.42	10.92 - 16.26	140	0.21
22-24	588	74	12.59	10.01 - 15.54	120	0.20
25-27	534	57	10.67	8.19 - 13.61	98	0.18
28-30	273	43	15.75	11.64 - 20.62	65	0.24

With respect to the percentage of offenders who blew over the 0.08 limit it started at 14.89% in the first three months and ended at 5.13%. This represents a reduction of 65.5% (100*(14.89-5.13)/14.89) of blows over 0.08. For the average blows over 0.08 per offender there is decreasing pattern although not as steady as for over 0.02 and 0.05. In this case it begins in 0.25 with subsequent decreases and increases (always below the 0.25) ending with 0.07 in the last period. This non-steady pattern might be due to the lower number of events so the data are more volatile toward the end of the tracking period.

Table 8-6: Participants who blew over 0.08						
Month	Participants in the program	Participants over limit	% Participants over limit	95%-CI	# fails over limit	Average fails/per Participant
1-3	1323	197	14.89	13.01 - 16.92	329	0.25
4-6	1298	128	9.86	8.29 - 11.61	196	0.15
7-9	1240	100	8.06	6.61 - 9.72	135	0.11
10-12	957	81	8.46	6.78 - 10.41	119	0.12
13-15	819	44	5.37	3.93 - 7.15	93	0.11
16-18	711	40	5.63	4.05 - 7.58	53	0.07
19-21	663	33	4.98	3.45 - 6.92	43	0.06
22-24	588	42	7.14	5.20 - 9.53	48	0.08
25-27	534	18	3.37	2.01 - 5.28	25	0.05
28-30	273	14	5.13	2.83 - 8.45	19	0.07

8.2.1 By gender

The tables in this subsection examine the behaviour of both females and males with respect to fails over 0.02, 0.05 and 0.08 respectively. However, note that the number of fails over 0.05 and 0.08 become too small over time (especially for females) for these results to be considered robust.

Regarding participants blowing over 0.02, males have a larger reduction of 34.9% (100*(69.86-45.45)/69.86) than females with 30% (100*(61.27-42.86)/61.27), but the male percentages are generally larger than the female percentages (except in the months 10-12). However, a test on the equality of proportions indicates that the equality of both percentages cannot be rejected (p-value=0.34). As for the average blows per offender, males exhibit a clear decreasing pattern beginning at 2.46 to 1 in months 25-27, followed by an increase to 1.36 in the last period. The average blows per female participant are usually below the average for males except for the periods between the 10th and the 15th month where the female average shows an increasing pattern.

Table 8-7: Percentage of participants who blew over 0.02: females						
Month	% females over limit	95%-CI			# fails over limit	Average fails/per females
1-3	61.27	53.26	-	69.28	301	2.12
4-6	60.28	52.21	-	68.36	227	1.61
7-9	48.53	40.13	-	56.93	149	1.10
10-12	57.78	47.57	-	67.98	150	1.67
13-15	45.07	33.50	-	56.64	144	2.03
16-18	40.30	28.55	-	52.04	77	1.15
19-21	41.27	29.11	-	53.43	73	1.16
22-24	40.00	26.42	-	53.58	32	0.64
25-27	26.19	12.89	-	39.49	30	0.71
28-30	42.86	21.69	-	64.02	22	1.05
Percentage of participants who blew over 0.02: males						
Month	% males over limit	95%-CI			# fails over limit	Average fails/per males
1-3	69.86	67.15	-	72.46	2909	2.46
4-6	63.61	60.77	-	66.39	2236	1.93
7-9	54.44	51.45	-	57.41	1718	1.56
10-12	52.70	49.32	-	56.07	1206	1.39
13-15	51.20	47.56	-	54.83	1008	1.34
16-18	49.61	45.69	-	53.54	852	1.32
19-21	50.33	46.26	-	54.40	788	1.31
22-24	46.85	42.58	-	51.16	681	1.26
25-27	36.11	31.86	-	40.52	495	1.00
28-30	45.45	39.21	-	51.81	344	1.36

Regarding participants blowing over 0.05, males have a larger reduction of 52.1% (100*(32.18-15.42)/32.18) than females with 33.9% (100*(30.28-20)/30.28). Although the percentages of change are significantly different (p-value<0.0001), the number of violations becomes very small over time (especially for females) for this result to be considered robust.

Table 8-8: Percentage of participants who blew over 0.05: females						
Month	% females over limit	95%-CI			# fails over limit	Average fails/per females
1-3	30.28	22.86	-	38.55	101	0.71
4-6	29.08	21.74	-	37.32	75	0.53
7-9	16.18	10.42	-	23.46	32	0.24
10-12	26.14	17.34	-	36.59	46	0.52
13-15	14.49	7.17	-	25.04	61	0.88
16-18	10.61	4.37	-	20.64	12	0.18
19-21	19.67	10.60	-	31.84	17	0.28
22-24	8.33	2.32	-	19.98	6	0.13
25-27	17.07	7.15	-	32.06	9	0.22
28-30	20.00	5.73	-	43.66	6	0.30

Percentage of participants who blew over 0.05: males						
Month	% males over limit	95%-CI			# fails over limit	Average fails/per males
1-3	32.18	29.52	-	34.92	751	0.64
4-6	25.24	22.76	-	27.84	491	0.42
7-9	18.66	16.40	-	21.09	335	0.30
10-12	16.23	13.83	-	18.85	263	0.30
13-15	16.13	13.57	-	18.96	187	0.25
16-18	15.19	12.51	-	18.20	169	0.26
19-21	12.79	10.23	-	15.73	123	0.20
22-24	12.96	10.25	-	16.09	114	0.21
25-27	10.14	7.62	-	13.15	89	0.18
28-30	15.42	11.20	-	20.46	59	0.23

Regarding participants blowing over 0.08, males have a larger reduction of 68% (100*(14.82-4.74)/14.82) than females with 35.44% (100*(15.49-10)/15.49). However, the number of violations becomes very small over time (especially for females) for this result to be considered robust.

Table 8-9: Percentage of participants who blew over 0.08: females					
Month	% females over limit	95%-CI			Average fails/per females
1-3	15.49	9.97	-	22.51	0.30
4-6	12.06	7.18	-	18.60	0.21
7-9	8.82	4.64	-	14.91	0.10
10-12	12.50	6.41	-	21.27	0.22
13-15	7.25	2.39	-	16.11	0.57
16-18	6.06	1.68	-	14.80	0.09
19-21	9.84	3.70	-	20.19	0.10
22-24	4.17	0.51	-	14.25	0.04
25-27	4.88	0.60	-	16.53	0.05
28-30	10.00	1.23	-	31.70	0.15
Percentage of participants who blew over 0.08: males					
Month	**% males over limit**	**95%-CI**			**Average fails/per males**
1-3	14.82	12.84	-	16.97	0.24
4-6	9.59	7.96	-	11.44	0.14
7-9	7.97	6.44	-	9.73	0.11
10-12	8.06	6.33	-	10.07	0.12
13-15	5.20	3.72	-	7.04	0.07
16-18	5.58	3.94	-	7.64	0.07
19-21	4.49	2.98	-	6.46	0.06
22-24	7.41	5.34	-	9.95	0.09
25-27	3.25	1.87	-	5.22	0.05
28-30	4.74	2.47	-	8.14	0.06

8.2.2 By mandatory versus voluntary status

Table 8-10, Table 8-11 and Table 8-12 examine the behaviour of both mandatory and voluntary groups with respect to fails over 0.02, 0.05 and 0.08 respectively. In the case of voluntary participants results are shown only for a period of 12 months since after that period only a small number remain in the program (remember this is the group of low risk that originally participates in the program for a period no longer than 12 months).

Regarding participants blowing over 0.02, voluntary participants have a larger reduction of 50.3% (100*(69.8-34.68)/69.8) than mandatory with 16.1% (100*(68.43-57.4)/68.43) in a 12 month period. In the 30 month period the mandatory group has a reduction of 34.3%. A test on the equality of proportions indicates that the equality of both percentages can be rejected at the 5% significance level (p-value<0.0001). As for the average blows per offender both groups show a clear decreasing pattern. In the case of the mandatory participants it starts at 2.69 and decreases to 0.98 in the months 25-27, followed by an increase to 1.33 in the last period. The average blows per voluntary participant are below the average for mandatory and it starts at 1.98 decreasing to 0.67 in the months 10-12.

Table 8-10: Percentage of participants who blew over 0.02 by status

Month	Mandatory in the program	Mandatory over limit	% mandatory over limit	95%-CI			# fails over limit	Average fails/per mandatory
1-3	833	570	68.43	65.15	-	71.57	2238	2.69
4-6	817	533	65.24	61.86	-	68.51	1778	2.18
7-9	804	476	59.20	55.72	-	62.62	1467	1.82
10-12	784	450	57.40	53.85	-	60.89	1240	1.58
13-15	763	394	51.64	48.03	-	55.24	1041	1.36
16-18	693	334	48.20	44.42	-	51.99	896	1.29
19-21	648	320	49.38	45.47	-	53.30	831	1.28
22-24	580	268	46.21	42.09	-	50.36	691	1.19
25-27	527	186	35.29	31.21	-	39.54	519	0.98
28-30	267	120	44.94	38.88	-	51.13	355	1.33
Month	**Voluntary in the program**	**voluntary over limit**	**% voluntary over limit**	**95%-CI**			**# fails over limit**	**Average fails/per voluntary**
1-3	490	342	69.80	65.52	-	73.83	972	1.98
4-6	481	288	59.88	55.34	-	64.29	685	1.42
7-9	436	191	43.81	39.09	-	48.61	400	0.92
10-12	173	60	34.68	27.62	-	42.28	116	0.67

Regarding participants blowing over 0.05, voluntary participants have a larger reduction of 71.9% 100*(26.73-7.51)/26.73) than mandatory participants with 45% (100*(35.05-19.26)/35.05) in a 12 month period. In the 30 month period the mandatory group has a reduction of 55.12% (100*(35.05-15.73)/35.05). As for the average blows per offender both groups show a clear decreasing pattern. In the case of the mandatory participants it starts at 0.77 and decreases to 0.18 in the months 25-27, followed by an increase to 0.24 in the last period. The average blows per voluntary participant are below the average for mandatory and it starts at 0.42 decreasing to 0.12.

Table 8-11: Percentage of participants who blew over 0.05 by status						
Month	**Mandatory in the program**	**Mandatory over limit**	**% mandatory over limit**	**95%-CI**	**# fails over limit**	**Average fails/per mandatory**
1-3	833	292	35.05	31.81 - 38.40	644	0.77
4-6	817	243	29.74	26.63 - 33.01	438	0.54
7-9	804	189	23.51	20.62 - 26.59	314	0.39
10-12	784	151	19.26	16.56 - 22.20	288	0.37
13-15	763	126	16.51	13.95 - 19.34	198	0.26
16-18	693	103	14.86	12.30 - 17.73	176	0.25
19-21	648	85	13.12	10.61 - 15.96	132	0.20
22-24	580	73	12.59	10.00 - 15.56	119	0.21
25-27	527	56	10.63	8.13 - 13.58	97	0.18
28-30	267	42	15.73	11.58 - 20.66	64	0.24
Month	**Voluntary in the program**	**voluntary over limit**	**% voluntary over limit**	**95%-CI**	**# fails over limit**	**Average fails/per voluntary**
1-3	490	131	26.73	22.86 - 30.89	208	0.42
4-6	481	90	18.71	15.32 - 22.49	128	0.27
7-9	436	39	8.94	6.44 - 12.03	53	0.12
10-12	173	13	7.51	4.06 - 12.51	21	0.12

Regarding participants blowing over 0.08, voluntary participants have a larger reduction of 55.6% reduction (100*(10.41-4.62)/10.41) than mandatory participants with 46.9% (100*(17.53-9.31)/17.53), in a 12 month period. However, the mandatory participants achieved a larger reduction of 70.1% (100*(17.53-5.24)/17.53) in a 30 month period. As for the average blows per offender both groups show a clear decreasing pattern. In the case of the mandatory participants it starts at 0.3 and decreases to 0.05 in months 25-27, followed by an increase to 0.07 in the last period. The average blows per voluntary participant starts at 0.16 decreasing to 0.08.

Table 8-12: Percentage of participants who blew over 0.08 by status						
Month	**Mandatory in the program**	**Mandatory over limit**	**% mandatory over limit**	**95%-CI**	**# fails over limit**	**Average fails/per mandatory**
1-3	833	146	17.53	15.00 - 20.28	252	0.30
4-6	817	106	12.97	10.75 - 15.47	165	0.20
7-9	804	87	10.82	8.76 - 13.18	119	0.15
10-12	784	73	9.31	7.37 - 11.57	106	0.14
13-15	763	42	5.50	4.00 - 7.37	57	0.07
16-18	693	38	5.48	3.91 - 7.45	51	0.07
19-21	648	31	4.78	3.27 - 6.72	41	0.06
22-24	580	42	7.24	5.27 - 9.66	48	0.08
25-27	527	18	3.42	2.04 - 5.34	25	0.05
28-30	267	14	5.24	2.90 - 8.64	19	0.07

Table 8-12: Percentage of participants who blew over 0.08 by status						
Month	Voluntary in the program	voluntary over limit	% voluntary over limit	95%-CI	# fails over limit	Average fails/per voluntary
1-3	490	51	10.41	7.85 - 13.46	77	0.16
4-6	481	22	4.57	2.89 - 6.84	31	0.06
7-9	436	13	2.98	1.60 - 5.04	16	0.04
10-12	173	8	4.62	2.02 - 8.91	13	0.08

8.2.3 By condition 37

Condition 37 is a condition on the driver's license requiring a zero BAC which is a discretionary decision of the Registrar of motor vehicles. Table 8-13, Table 8-14 and Table 8-15 examine the behaviour of participants with this condition on their license versus those without the condition, with respect to fails over 0.02, 0.05 and 0.08 respectively. However, note that the number of fails over 0.05 and 0.08 become too small over time (especially for participants without condition 37) for these results to be considered robust.

Regarding participants blowing over 0.02, participants without the condition have a larger reduction of 57% (100*(65.08-28)/65.08) than participants with the condition with 30% (100*(72.93-51.05)/72.93). As for the average blows per offender both groups show a clear decreasing pattern. In the case of the participants with the condition it starts at 3 and decreases to 1.25 in the months 25-27, followed by an increase to 1.5 in the last period. The average blows per participant without the condition are below the average for those with the condition and it starts at 1.83 decreasing to 0.73.

Table 8-13: Percentage of participants who blew over 0.02 by condition 37						
Month	Participants with cond37 in the program	Participants with cond37 over limit	% Participants with cond37 over limit	95%-CI	# fails over limit	Average fails/per Participants with cond37
1-3	639	466	72.93	69.30 - 76.34	1934	3.03
4-6	628	423	67.36	63.53 - 71.01	1493	2.38
7-9	613	369	60.20	56.20 - 64.10	1148	1.87
10-12	555	342	61.62	57.43 - 65.69	982	1.77
13-15	518	307	59.27	54.90 - 63.53	905	1.75
16-18	462	257	55.63	50.97 - 60.22	745	1.61
19-21	433	245	56.58	51.77 - 61.31	680	1.57
22-24	368	187	50.82	45.58 - 56.03	552	1.50
25-27	322	132	40.99	35.57 - 46.58	402	1.25
28-30	190	97	51.05	43.71 - 58.36	285	1.50

Table 8-13: Percentage of participants who blew over 0.02 by condition 37								
Month	Participants without cond37 in the program	Participants without cond37 over limit	% Participants without cond37 over limit	95%-CI			# fails over limit	Average fails/per Participants without cond37
1-3	650	423	65.08	61.27	-	68.74	1187	1.83
4-6	637	376	59.03	55.09	-	62.87	908	1.43
7-9	599	283	47.25	43.19	-	51.33	673	1.12
10-12	377	153	40.58	35.59	-	45.73	340	0.90
13-15	280	101	36.07	30.44	-	42.00	220	0.79
16-18	233	83	35.62	29.48	-	42.14	163	0.70
19-21	216	74	34.26	27.95	-	41.00	156	0.72
22-24	207	77	37.20	30.60	-	44.17	135	0.65
25-27	199	50	25.13	19.26	-	31.75	110	0.55
28-30	75	21	28.00	18.24	-	39.56	55	0.73

Table 8-14: Percentage of participants who blew over 0.05 by condition 37								
Month	Participants with cond37 in the program	Participants with cond37 over limit	% Participants with cond37 over limit	95%-CI			# fails over limit	Average fails/per Participants with cond37
1-3	639	246	38.50	34.71	-	42.39	558	0.87
4-6	628	204	32.48	28.83	-	36.30	384	0.61
7-9	613	148	24.14	20.81	-	27.73	240	0.39
10-12	555	125	22.52	19.11	-	26.23	255	0.46
13-15	518	99	19.11	15.81	-	22.77	209	0.40
16-18	462	87	18.83	15.37	-	22.70	155	0.34
19-21	433	70	16.17	12.82	-	19.98	116	0.27
22-24	368	60	16.30	12.68	-	20.48	100	0.27
25-27	322	43	13.35	9.84	-	17.56	83	0.26
28-30	190	35	18.42	13.18	-	24.68	53	0.28
Month	Participants without cond37 in the program	Participants without cond37 over limit	% Participants without cond37 over limit	95%-CI			# fails over limit	Average fails/per Participants without cond37
1-3	650	163	25.08	21.79	-	28.59	275	0.42
4-6	637	116	18.21	15.29	-	21.43	159	0.25
7-9	599	75	12.52	9.98	-	15.44	117	0.20
10-12	377	36	9.55	6.78	-	12.97	49	0.13
13-15	280	28	10.00	6.75	-	14.13	33	0.12
16-18	233	17	7.30	4.31	-	11.42	21	0.09

Table 8-14: Percentage of participants who blew over 0.05 by condition 37								
19-21	216	14	6.48	3.59	-	10.64	19	0.09
22-24	207	11	5.31	2.68	-	9.31	16	0.08
25-27	199	13	6.53	3.52	-	10.91	14	0.07
28-30	75	5	6.67	2.20	-	14.88	6	0.08

Table 8-15: Percentage of participants who blew over 0.08 by condition 37

Month	Participants with cond37 in the program	Participants with cond37 over limit	% Participants with cond37 over limit	95%-CI			# fails over limit	Average fails/per Participants with cond37
1-3	639	116	18.15	15.24	-	21.37	215	0.34
4-6	628	86	13.69	11.10	-	16.63	142	0.23
7-9	613	62	10.11	7.84	-	12.78	82	0.13
10-12	555	60	10.81	8.35	-	13.70	96	0.17
13-15	518	34	6.56	4.59	-	9.05	82	0.16
16-18	462	32	6.93	4.79	-	9.64	45	0.10
19-21	433	26	6.00	3.96	-	8.67	36	0.08
22-24	368	34	9.24	6.48	-	12.67	40	0.11
25-27	322	15	4.66	2.63	-	7.57	22	0.07
28-30	190	10	5.26	2.55	-	9.47	13	0.07
Month	Participants without cond37 in the program	Participants without cond37 over limit	% Participants without cond37 over limit	95%-CI			# fails over limit	Average fails/per Participants without cond37
1-3	650	76	11.69	9.32	-	14.42	108	0.17
4-6	637	38	5.97	4.26	-	8.10	49	0.08
7-9	599	34	5.68	3.96	-	7.84	47	0.08
10-12	377	19	5.04	3.06	-	7.76	21	0.06
13-15	280	8	2.86	1.24	-	5.55	9	0.03
16-18	233	7	3.00	1.22	-	6.09	7	0.03
19-21	216	6	2.78	1.03	-	5.95	6	0.03
22-24	207	6	2.90	1.07	-	6.20	6	0.03
25-27	199	3	1.51	0.31	-	4.34	3	0.02
28-30	75	1	1.33	0.03	-	7.21	1	0.01

Regarding participants blowing over 0.05, participants without the condition have a larger reduction of 73% (100 (25.08-6.67)/25.08) than participants with the condition with 52% (100*(38.5-18.42)/38.5). A test on the equality of proportions indicates that the equality of both percentages can be rejected at the 5% significance level (p-value<0.0001). Thus, there is strong evidence that the participants without the condition have a larger reduction in the percentage failing over the 0.05 limit. As for the average blows per offender both groups show a clear decreasing pattern. In the case of the participants with the condition it starts at 0.87 and decreases to 0.28. The average blows per participant without the condition are below the average for those with the condition and it starts at 0.42 decreasing to 0.08.

For percent of participants blowing over 0.08, participants without the condition have a larger reduction of 88% (11.69-1.33)/11.69) than participants with the condition with 71% (100*(18.15-5.26)/18.15). A test on the equality of proportions indicates that the equality of both percentages can be rejected at the 5% significance level (p-value<0.0001). Again, there is strong evidence that the participants without the condition have a larger reduction in the percentage failing over the 0.08 limit. As for the average blows per offender both groups show a clear decreasing pattern. In case of the participants with the condition it starts at 0.34 and decreases to 0.07. The average blows per participant without the condition are below the average for those with the condition and it starts at 0.17 decreasing to 0.01.

8.2.4 By device

According to the data, the participants in the interlock program used two types of devices: the WR3 and the LR. Table 8-16 examines the behaviour of participants with the WR3 device versus those with the LR device, with respect to fails over 0.02.

Table 8-16: Percentage of participants who blew over 0.02 by device

Month	% Participants with WR3 over limit	95%-CI	# fails over limit	Average fails/per Participants with WR3
1-3	68.93	66.36 - 71.42	3210	2.43
4-6	63.25	60.56 - 65.88	2463	1.90
7-9	53.79	50.97 - 56.59	1867	1.51
10-12	53.29	50.07 - 56.49	1354	1.41
13-15	50.61	47.12 - 54.10	1142	1.40
16-18	48.57	44.81 - 52.34	900	1.29
19-21	49.31	45.40 - 53.22	840	1.29
22-24	46.19	42.07 - 50.36	686	1.19
25-27	34.67	30.60 - 38.91	480	0.91
28-30	44.61	38.57 - 50.77	352	1.31
Month	**% Participants with LR over limit**	**95%-CI**	**# fails over limit**	**Average fails/per Participants with LR**
1-3	0		0	0
4-6	0		0	0
7-9	0		0	0
10-12	0		0	0
13-15	80.00	28.36 - 99.49	10	2.00
16-18	63.64	30.79 - 89.07	29	2.64
19-21	66.67	34.89 - 90.08	21	1.75
22-24	60.00	26.24 - 87.84	27	2.70
25-27	77.78	39.99 - 97.19	45	5.00
28-30	100.00	39.76 - 100.00	14	3.50

However, during the study period the LR device was used for the first time in August 27, 2013, when the drivers in our study were past their ninth month in the program (participants in the study started in the program up to December 2012). Due to the limited data with the LR device it is not possible to properly compare the effect of the type of device on the participants' behaviour.

8.2.5 Start-up versus running retests fails

Table 8-17 and Table 8-18 examine the behaviour of participants at start-up and running retests respectively. In general the number of fails at start-up is larger than the fails at running retests. The percent of participants failing at start-up have a larger reduction of 36.7% (100*(60.17-38.10)/60.17) than participants failing at running retests with 24.6% (100*(28.19-21.25)/28.19). As for the average fails per offender, both type of events show a decreasing pattern. In the case of the fails at start-up it starts at 1.97 and decreases to 0.76 in months 25-27, followed by an increase to 1.05 in the last period. The average fails per participant at running retests starts at 0.46 decreasing to 0.29.

Table 8-17: Fails at start-up

Month	Participants in the program	Participants over limit	% Participants over limit	95%-CI			# fails over limit	Average fails/per Participant
1-3	1323	796	60.17	57.47	-	62.82	2606	1.97
4-6	1298	708	54.55	51.79	-	57.28	1944	1.50
7-9	1240	535	43.15	40.37	-	45.96	1412	1.14
10-12	957	419	43.78	40.61	-	46.99	1069	1.12
13-15	819	347	42.37	38.96	-	45.84	907	1.11
16-18	711	288	40.51	36.87	-	44.22	735	1.03
19-21	663	262	39.52	35.77	-	43.35	640	0.97
22-24	588	210	35.71	31.84	-	39.74	535	0.91
25-27	534	153	28.65	24.85	-	32.69	406	0.76
28-30	273	104	38.10	32.31	-	44.14	286	1.05

Table 8-18: Overall fails at running retests

Month	Participants in the program	Participants over limit	% Participants over limit	95%-CI			# fails over limit	Average fails/per Participant
1-3	1323	373	28.19	25.78	-	30.70	604	0.46
4-6	1298	343	26.43	24.04	-	28.91	519	0.40
7-9	1240	296	23.87	21.52	-	26.34	455	0.37
10-12	957	197	20.59	18.06	-	23.29	287	0.30
13-15	819	160	19.54	16.87	-	22.42	245	0.30
16-18	711	139	19.55	16.70	-	22.66	194	0.27
19-21	663	153	23.08	19.92	-	26.48	221	0.33
22-24	588	127	21.60	18.34	-	25.15	178	0.30

Table 8-18: Overall fails at running retests						
Month	Participants in the program	Participants over limit	% Participants over limit	95%-CI	# fails over limit	Average fails/per Participant
25-27	534	80	14.98	12.06 - 18.29	119	0.22
28-30	273	58	21.25	16.55 - 26.58	80	0.29

8.2.6 Conclusions

Table 8-19 summarizes the previous results in terms of the overall change over time in the percentage of participants that fail tests according to each of the study factors (gender, status, condition 37, start-up/running). The results show that overall there is a reduction in the percentage of participants failing tests over time illustrating a learning effect. The effect is more pronounced in the more risky behaviour (fails over 0.08) with larger changes over the same time.

Although these results indicate that the learning effect is more pronounced in male participants than female participants, the amount of females with violations over 0.05 and 0.08 is too small rendering these results volatile. The difference with respect to the reduction in fails over 0.02 is not significantly different between both genders.

Table 8-19: Reductions over time in the percentage of participants with failed tests associated with each factor				
factor	time (months)	over 0.02	over 0.05	over 0.08
overall	30	34.1%	50.8%	65.5%
female	30	30.0%	33.9%	35.4%
male	30	34.9%	52.1%	68.0%
voluntary	12	50.3%	71.9%	55.6%
mandatory	12	16.1%	45.0%	46.9%
mandatory	30	34.3%	55.1%	70.1%
condition 37	30	30.0%	52.0%	71.0%
no-condition 37	30	57.0%	73.0%	88.0%
start-up tests	30	36.7%		
running retests	30	24.6%		

The learning effect is more pronounced in the voluntary participants than in the mandatory participants as they learned faster to become compliant over the same period of time. Nevertheless, the mandatory participants did also show a learning effect, and in a longer period of time they achieve larger changes, particularly with respect to fails over 0.08. Similarly, those without condition 37 show a more pronounced learning effect than the participants with the condition.

Finally, although the percentages of participants failing are larger at start-up tests than running retests, the learning effect is more pronounced (larger percentage reduction over time) with respect to start-up tests.

8.3 Fails in monthly periods

The previous analyses broke down the interlock events in periods of three months. In this section we present figures for similar analyses but considering the periods monthly instead of every three months. Figure 8-2 shows the number of participants, fails and participants failing over time by the different BAC limits (0.02, 0.05 and 0.08). The curves show a decline in the number of fails over time, but the number of participants also decreases over time.

Figure 8-2: Number of fails and participants failing over time

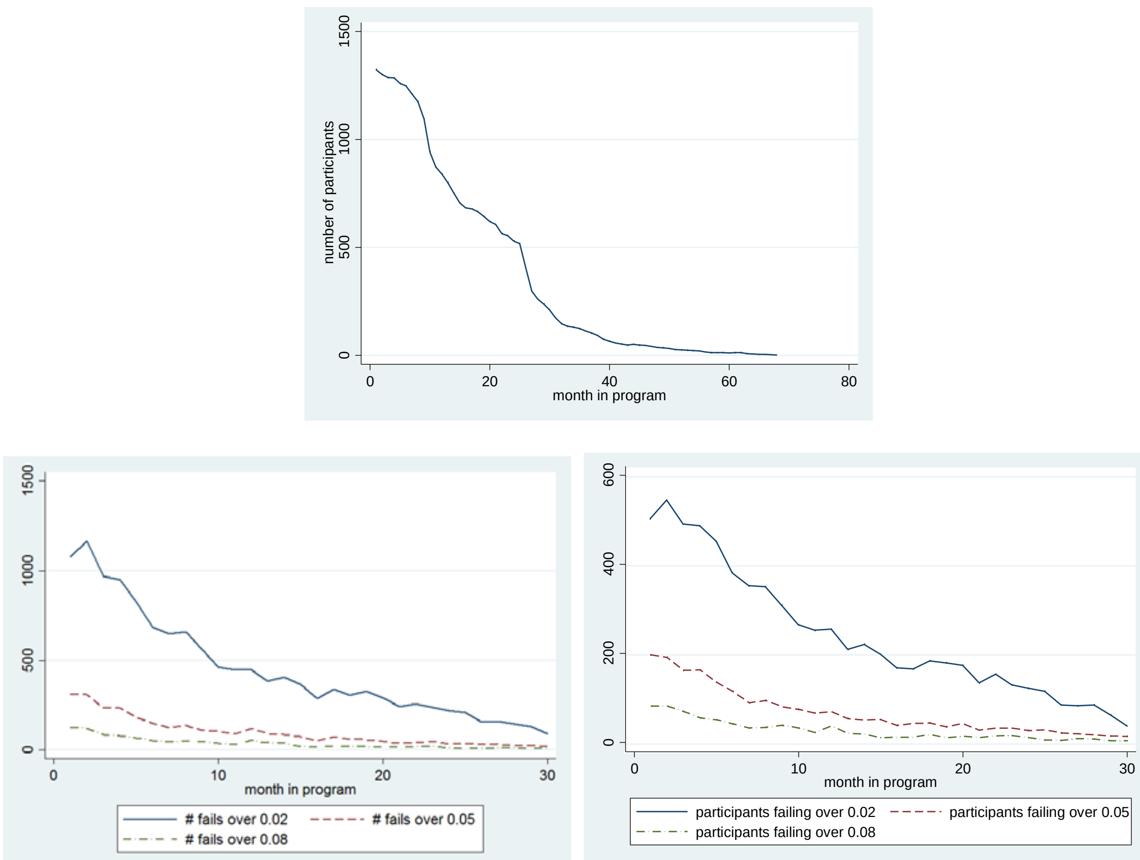
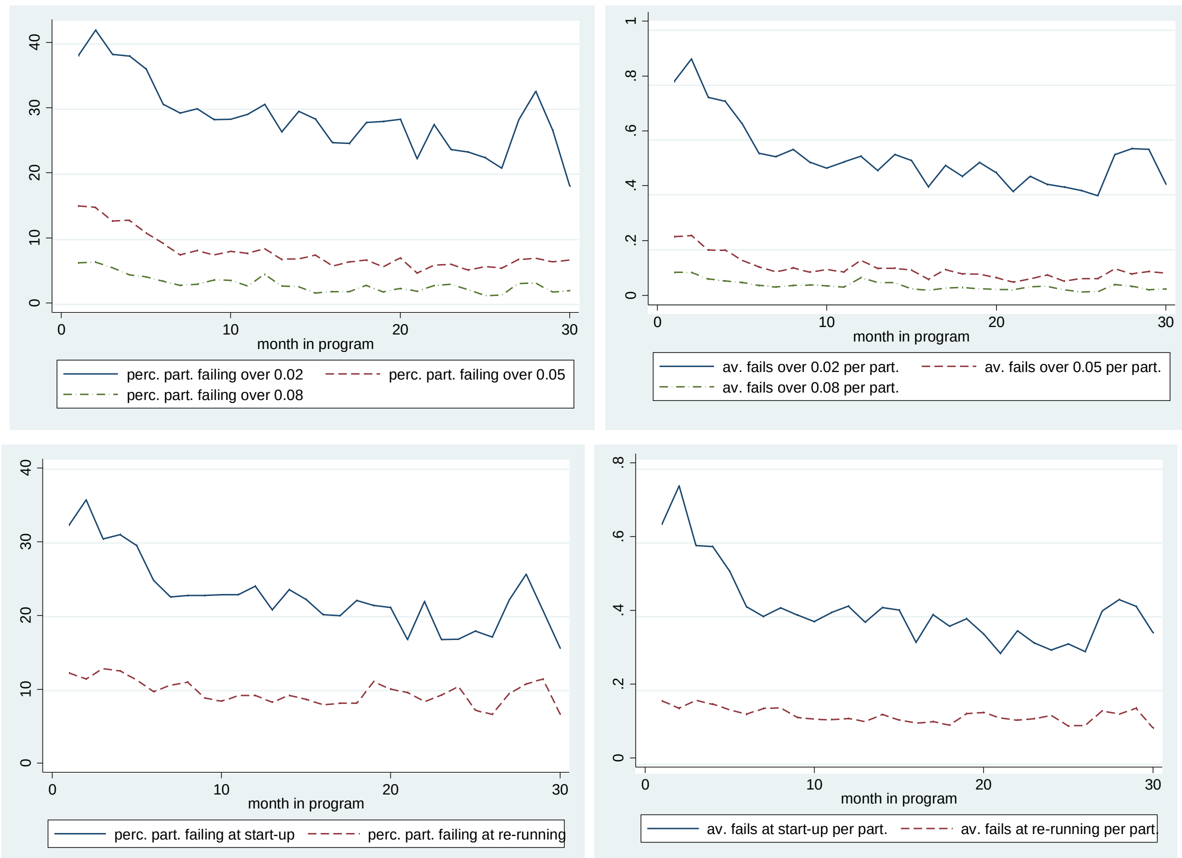


Figure 8-3 shows the percentage of participants failing and the average of fails by participants over time. The curves show that the percentage of participants failing and the average fails per participant decrease over time. The decrease is more pronounced in the first 10 months and then it becomes less pronounced and almost negligible in the subsequent months. Particularly interesting is that after the 24 month period the curves show an increasing pattern, more pronounced in the case of the less risky behaviour (BAC over 0.002).

Figure 8-3: Percentage of participants failing and average fails by participants



8.3.1 By gender

Figure 8-4 shows the number of participants by gender over time. Figure 8-5 shows the percentage of participants failing and the average of fails by participants over time for the different BAC limits (0.02, 0.05 and 0.08) by gender. Figure 8-6 shows the percentage of participants failing and the average of fails by participants over time at start-up and running retests by gender.

Figure 8-4: Number of participants by gender

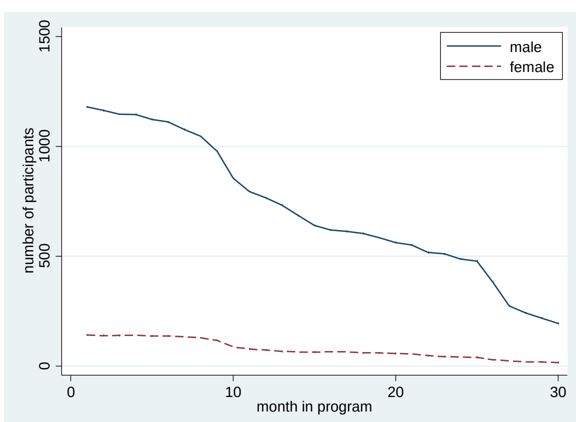


Figure 8-5: Percentage of participants failing and average fails by participants, by gender

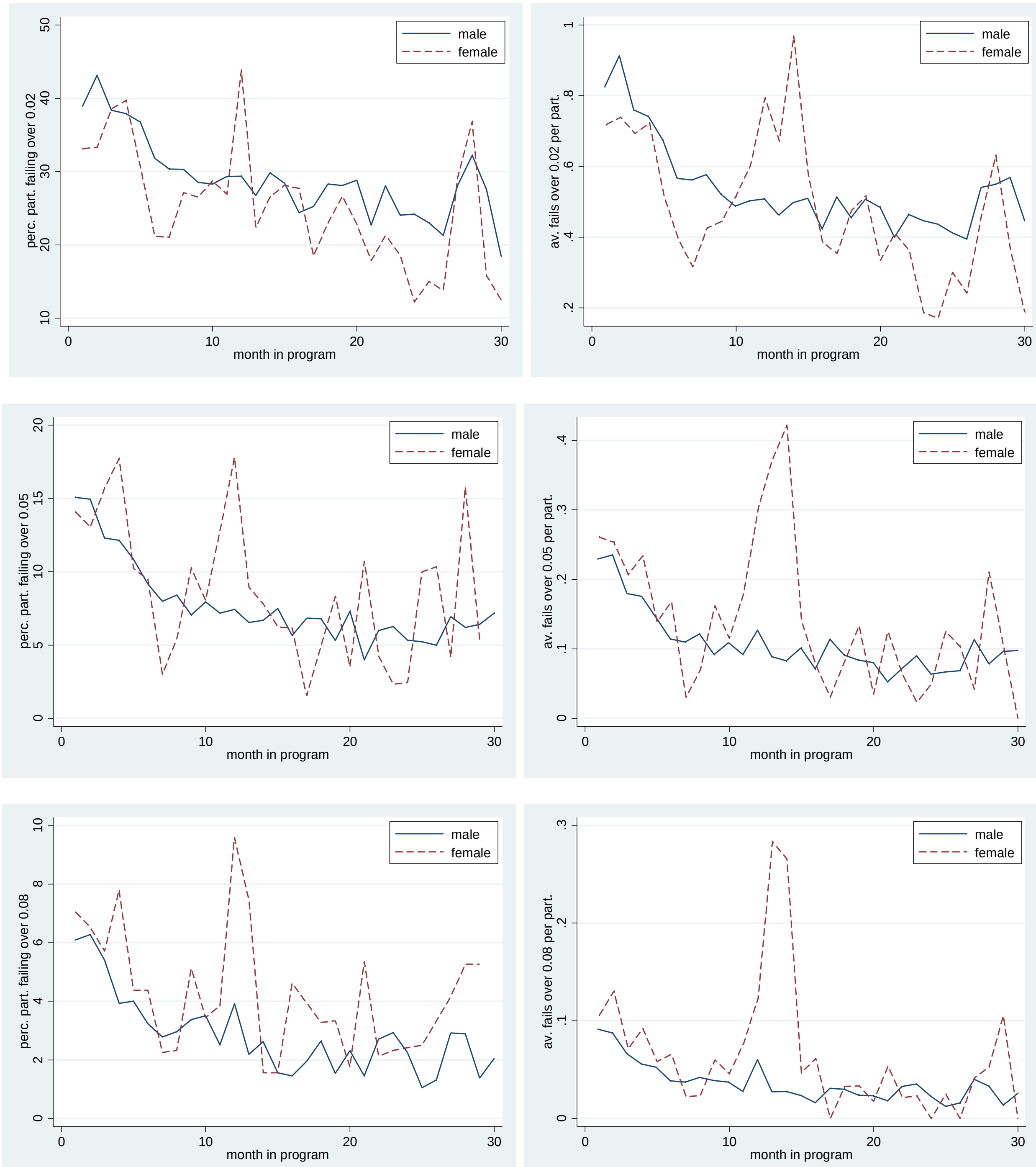
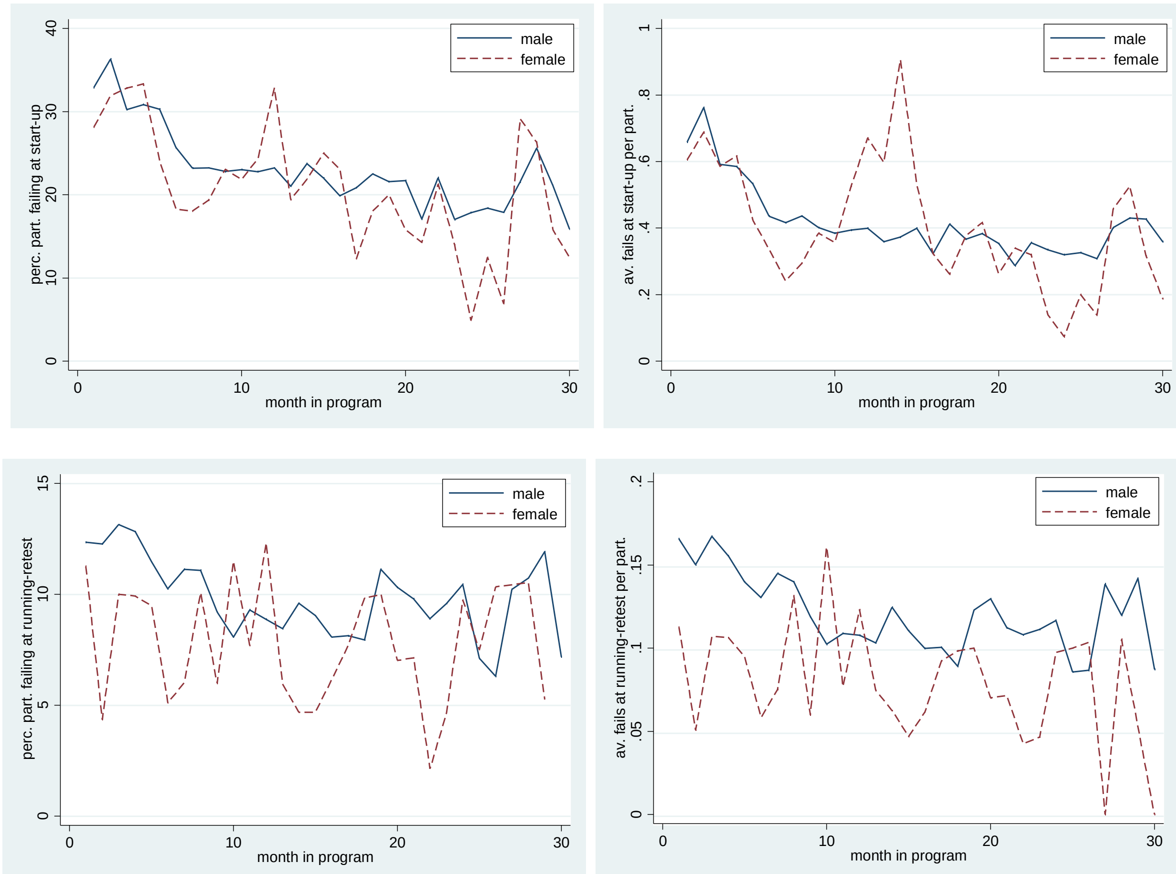


Figure 8-6: Percentage of participants failing and average fails by participants, by gender at start-up and running retests



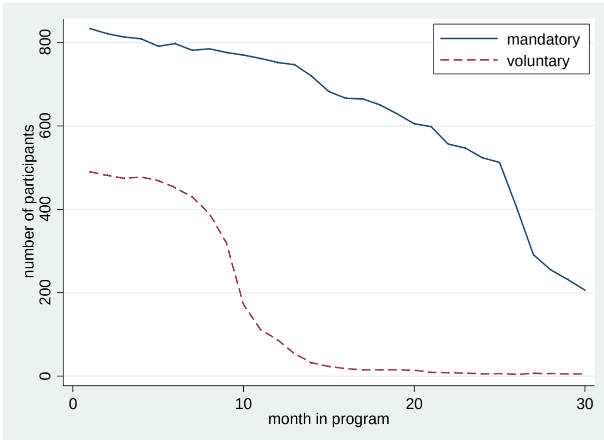
In general, the curves for males are similar to the overall curves in the previous section (males represent approximately 91% of the total participants) and they show a general decreasing pattern for the percentage of participants failing and the average fails per participant. On the other hand, the curves for females show irregular patterns, particularly after the tenth month in the program. In general, there seems to be a decreasing pattern in the curves for female up to about month ten, except for running retests.

Comparing female versus male participants there seems to be no large differences. Up to month ten in the program, the percentages and averages for failed tests over 0.02, at start-up and at running retests, seem smaller for females than for males. However, for failed tests over 0.05 and over 0.08, the curves for females oscillate around the curves for males with a prevalence of larger values for females. The more volatile nature of the data for females is related to the smaller number of female participants in the program, especially after ten months.

8.3.2 By mandatory versus voluntary

Figure 8-7 shows the number of participants by status over time. Although the voluntary participants were defined by those expected to be in the program for less than a year, some may have received an extension beyond a year.

Figure 8-7: Number of participants by status mandatory/voluntary



Since the number of voluntary participants beyond one year is very small, the next figures show data over a 12 month period only. Figure 8-8 shows the percentage of participants failing and the average of fails by participants over time for the different BAC limits (0.02, 0.05 and 0.08) by status. Figure 8-9 shows the percentage of participants failing and the average of fails by participants over time at start-up and running retest by status.

In general, the percentages and averages are smaller for the voluntary participants than for the mandatory ones. However, both mandatory and voluntary participants reveal a decreasing pattern in the percentage of participants failing and in the average fails per participant over time.

Figure 8-8: Percentage of participants failing and average fails by participants. By status

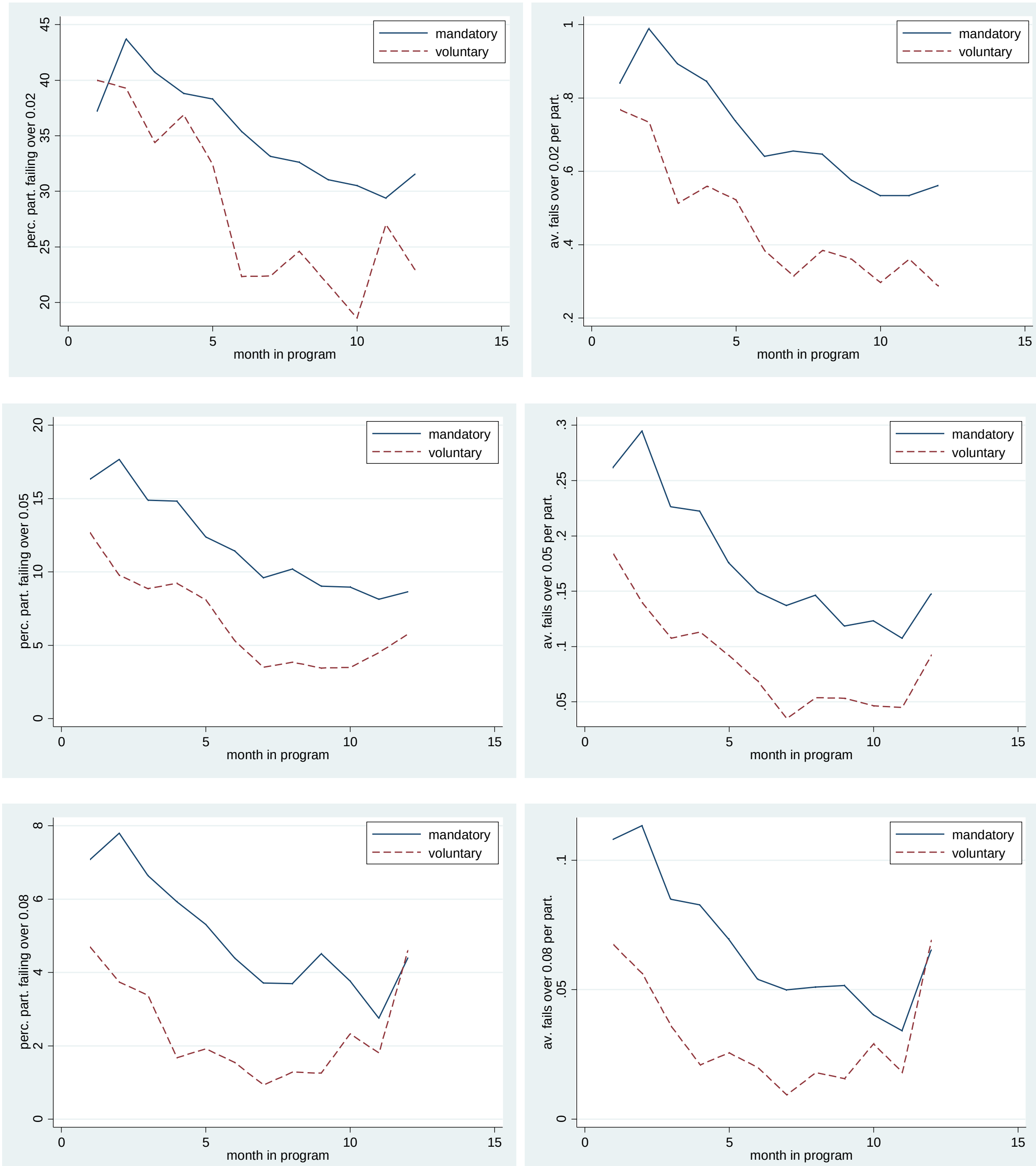
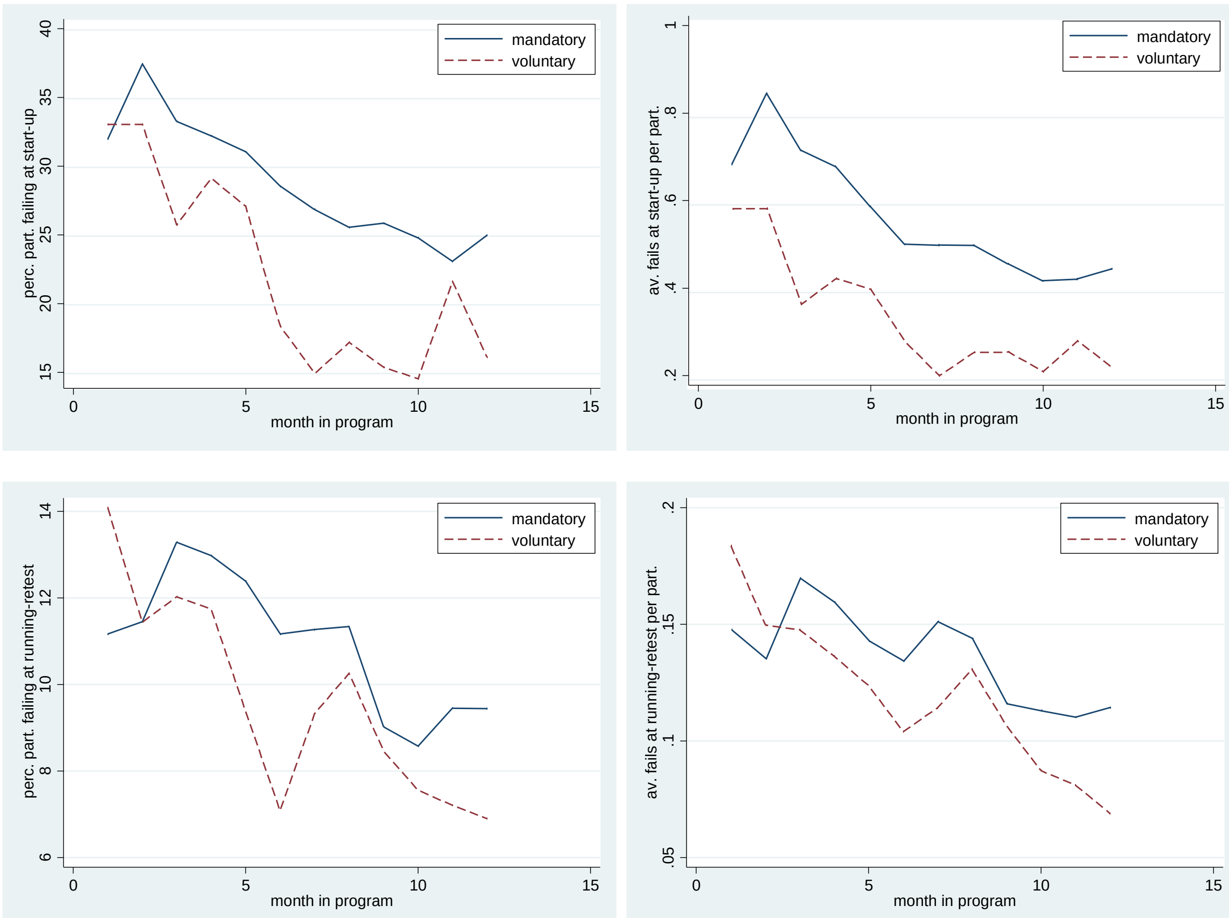


Figure 8-9: Percentage of participants failing and average fails by participants. By status at start-up and running



8.3.3 By condition 37

Figure 8-10 shows the number of participants by condition 37 over time.

Figure 8-10: Number of participants by condition 37

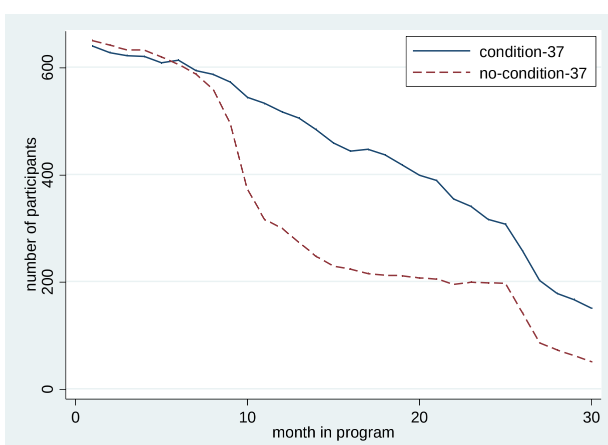


Figure 8-11 shows the percentage of participants failing and the average of fails by participants over time for the different BAC limits (0.02, 0.05 and 0.08) by condition 37. Figure 8-12 shows the percentage of participants failing and the average of fails by participants over time at start-up and running retest by condition 37.

Figure 8-11: Percentage of participants failing and average fails by participants. By condition 37

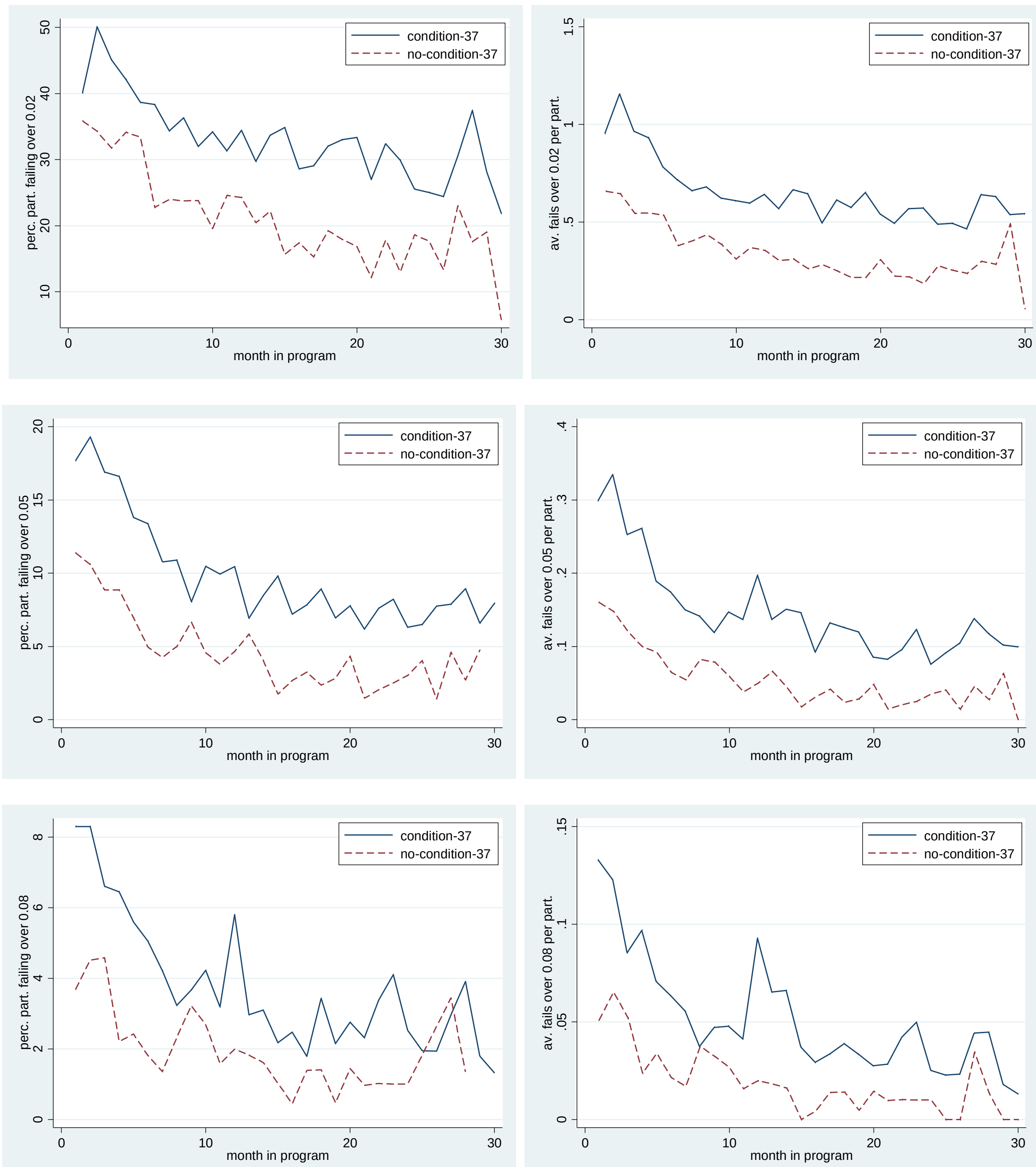
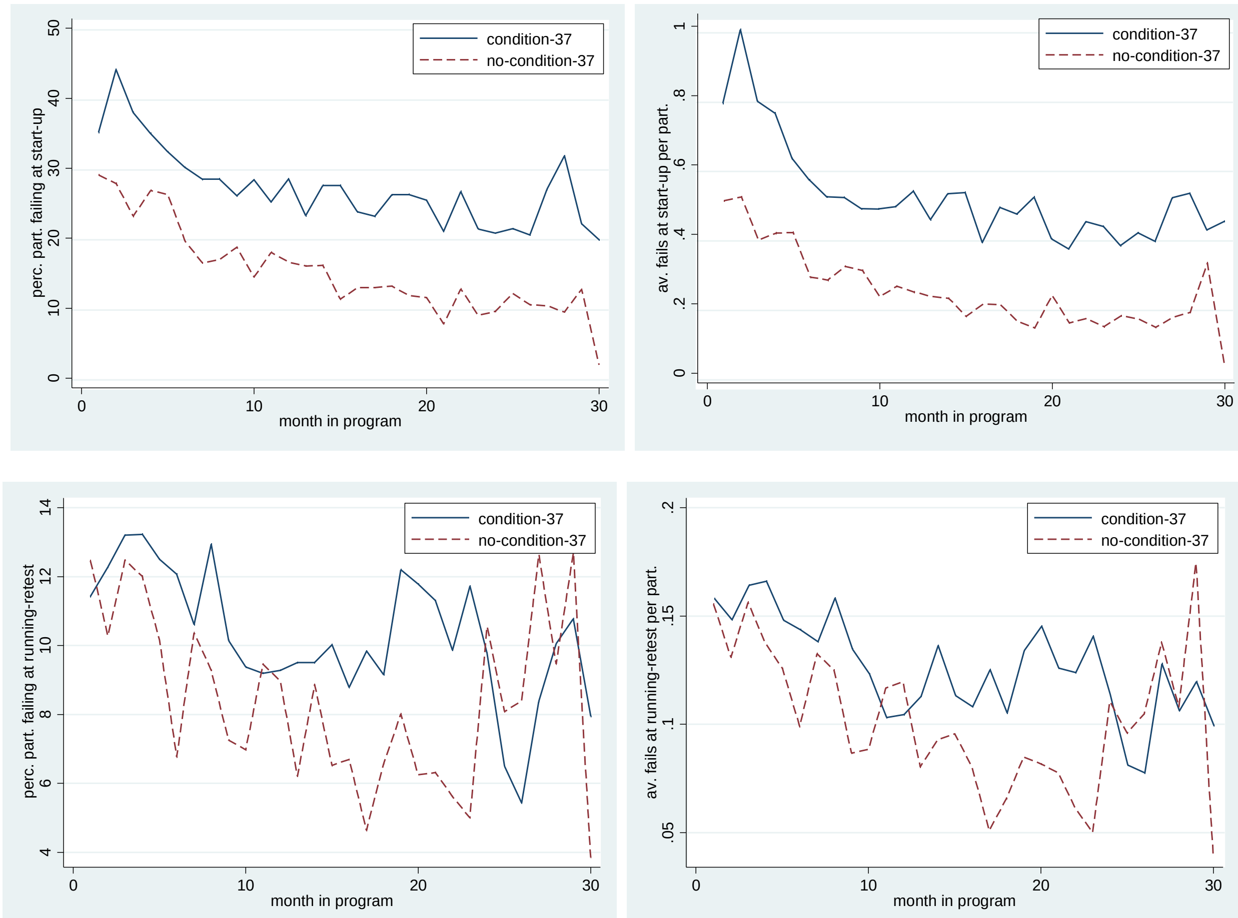


Figure 8-12: Percentage of participants failing and average fails by participants. By condition 37 at start-up and running



In general, the percentages and averages are smaller for the participants without the condition than for the participants with the condition. However, both groups of participants reveal a decreasing pattern in the percentage of participants failing and in the average fails per participant over time.

8.4 Logistic regression analysis

Logistic regression analysis enabled to compare the odds of a failed test to the odds of a passed test while simultaneously controlling for several factors like months in the program, gender, age, mandatory/voluntary status, start-up/running type of test, condition 37 and average mileage driven per month. The variable "mileagepm" represents the average in thousands of kilometers driven per month.

Figure 8-13 shows the estimated logistic model for the fails over the 0.02 limit in a 12 month tracking period. All estimated parameters are significant (p-values<0.05) except for gender and one age category, age 35-44 (p-values= 0.056 and 0.3). The results show that the odds for failing over the 0.02 limit:

- decrease over time (OR=0.94), 6% per month,
- for a mandatory participant are larger (OR=1.3) than for a voluntary participant,
- for a participant with condition 37 are larger (OR=1.5) than for a participant without the condition,
- at start-up are larger (OR=3.4) than at running retests,
- decrease with mileage driven (OR=0.97), 3% per 1000 kilometers.

Figure 8-13: Logistic model for fails over 0.02

```
Logistic regression                                  Number of obs   =    3942047
                                                     LR chi2(10)     =    3562.84
                                                     Prob > chi2     =     0.0000
Log likelihood = -59905.264                          Pseudo R2       =     0.0289
------------------------------------------------------------------------------
      over02 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       month |   .9395317   .0031298   -18.72   0.000     .9334173    .9456861
      gender |   .9299366   .0353009    -1.91   0.056     .8632592    1.001764
      agecat |
      25-34  |   .7332519   .0389018    -5.85   0.000     .6608359    .8136033
      35-44  |   .9320938   .0481067    -1.36   0.173     .8424186    1.031315
      45-64  |   .8510554   .0430194    -3.19   0.001      .770781    .9396901
    65 & over|   .7881751   .0553906    -3.39   0.001     .6867566    .9045708
             |
        mand |   1.291417   .0357425     9.24   0.000     1.223229    1.363406
      cond37 |   1.512322   .0368317    16.98   0.000     1.441829    1.586261
     startup |   3.417363   .0904074    46.45   0.000     3.244684    3.599233
    recmilpm |   .9740198   .0063464    -4.04   0.000     .9616601    .9865383
       _cons |   .0013292   .0001223   -71.96   0.000     .0011098     .001592
```

Figure 8-14 and Figure 8-15 show the estimated logistic model for fails over the 0.05 and 0.08 limits respectively. The estimated models are similar in the sense that the same factors are significant and the value of their odds ratios indicate that those same factors are more likely to be associated with fails in both models. However, the effect of the different factors seems to be slightly more pronounced for the more risky behaviour since the values of the odds ratios are more distant from 1.

Figure 8-14: Logistic model for fails over 0.05

Logistic regression — Number of obs = 3942047; LR chi2(10) = 2347.41; Prob > chi2 = 0.0000; Log likelihood = -16278.741; Pseudo R2 = 0.0673

over05	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
month	.909155	.0063161	-13.71	0.000	.8968596	.9216191
gender	.686917	.0474664	-5.43	0.000	.5999094	.7865436
agecat						
25-34	.6971245	.0797078	-3.16	0.002	.5571674	.8722381
35-44	.9275967	.1031214	-0.68	0.499	.745986	1.15342
45-64	.9227415	.1002325	-0.74	0.459	.7457935	1.141672
65 & over	.4758591	.0803137	-4.40	0.000	.3418344	.6624316
mand	1.699578	.1043051	8.64	0.000	1.506961	1.916816
cond37	1.88717	.0990036	12.11	0.000	1.702769	2.09154
startup	14.34605	1.350287	28.30	0.000	11.92931	17.25241
recmilpm	.921847	.0129603	-5.79	0.000	.8967921	.947602
_cons	.0001523	.0000297	-45.08	0.000	.0001039	.0002231

Figure 8-15: Logistic model for fails over 0.08

Logistic regression — Number of obs = 3942047; LR chi2(10) = 998.42; Prob > chi2 = 0.0000; Log likelihood = -6761.6729; Pseudo R2 = 0.0688

over08	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
month	.9060871	.010303	-8.67	0.000	.8861171	.9265073
gender	.627525	.0677619	-4.32	0.000	.5078272	.7754363
agecat						
25-34	.7893818	.1667264	-1.12	0.263	.5217996	1.194182
35-44	1.119309	.2297578	0.55	0.583	.748559	1.673686
45-64	1.270543	.254709	1.19	0.232	.8577224	1.882053
65 & over	.6360747	.1797213	-1.60	0.109	.3655975	1.106657
mand	2.013349	.2088836	6.75	0.000	1.642886	2.467349
cond37	1.798871	.1538463	6.87	0.000	1.521256	2.12715
startup	18.4061	3.185897	16.83	0.000	13.11076	25.84018
recmilpm	.8648253	.0211529	-5.94	0.000	.8243443	.9072941
_cons	.0000463	.0000156	-29.64	0.000	.000024	.0000897

One particular difference is with respect to gender. In the models for fails over the 0.05 and 0.08 limits, gender is significant (p-values<0.001). It should be noted however, that the differences in gender are actually small (0.07% versus 0.05% of fails or 0.03% versus 0.02%), see frequency for fails in Figure 8 – 16 and Figure 8 – 17. Due to the large sample size (3,942,047 observations) small effects are more easily found to be significant, even if the actual differences may not be meaningful.

Figure 8-16: Frequencies of fails over the 0.05 limit per gender

fails over0.05	gender F	M	Total
no	338,958 99.93	3,601,052 99.95	3,940,010 99.95
yes	245 0.07	1,792 0.05	2,037 0.05
Total	339,203 100.00	3,602,844 100.00	3,942,047 100.00

Pearson chi2(1) = 30.3602 Pr = 0.000

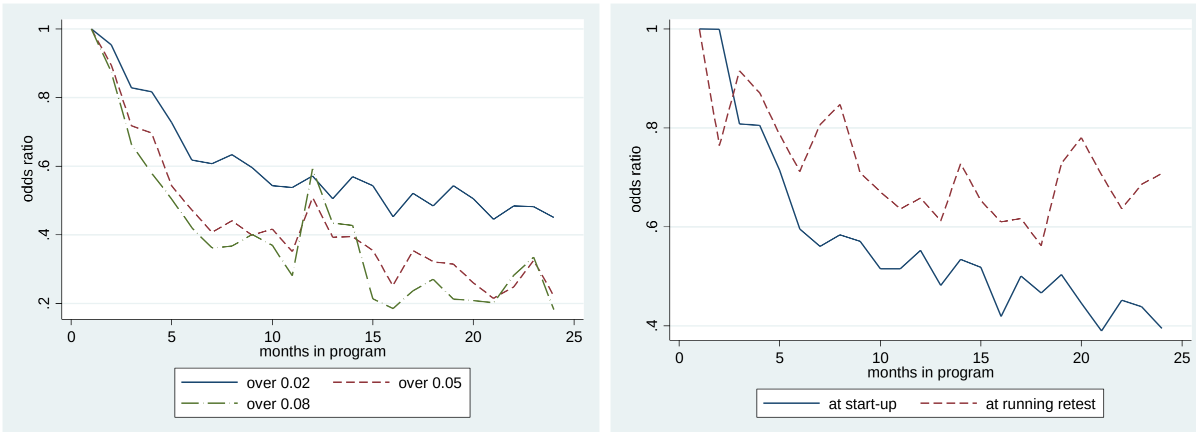
Figure 8-17: Frequencies of fails over the 0.08 limit per gender

fails over0.08	gender F	M	Total
no	339,101 99.97	3,602,186 99.98	3,941,287 99.98
yes	102 0.03	658 0.02	760 0.02
Total	339,203 100.00	3,602,844 100.00	3,942,047 100.00

Pearson chi2(1) = 22.4215 Pr = 0.000

With the logistic regression analysis it is possible to obtain the odds ratios for the specific values of categorical variables. This allows for example to obtain the odds ratios for each value of the month variable with respect to the first month and see how the odds for failing change over time from month to month. Figure 8-18 shows the change over time of the odds ratios for failed tests comparing the three limits (0.02, 0.05 and 0.08) and start-up fails versus running retest fails. The figure shows that the odds for failing decrease faster for the more risky behaviours (fail over 0.05 and 0.08) suggesting that participants learn faster to be compliant for these more risky behaviours. The comparison between failing at start-up and running retests suggests that participants learn to be compliant faster at start-up than at running retests.

Figure 8-18: Change over time of odds ratio for failed tests relative to month 1



8.5 Conclusions

The descriptive analysis in tables (three months period) and figures (monthly period) studied the percentage of participants failing and the average fails per participant over time. Different tables and figures are presented for the different studied factors (such as gender, status and condition 37). The logistic regression analysis compared the odds of a failed test to the odds of a passed test while simultaneously controlling for multiple factors. Different logistic models are presented for each type of failed test.

From the interlock data, overall the results suggests that there are “learning curves” illustrating that offenders were more likely to violate at the beginning of program participation (with larger percentage and average of violations per participants and larger odds for failing tests), but over time these violations decreased as offenders may learn about, or experience the consequences of, program violations and the nuances associated with the functioning of, and compliance with, devices. In general, the curves are steepest at the beginning of program participation until approximately month 10, indicating that the “learning effect” may decrease or stop after a period of time. However, it warrants mentioning that the decreasing pattern of the events we studied is not always a smooth one. Sometimes the curves reveal peaks, despite an overall decrease from the beginning until the end of the program, but it is not clear if those peaks are the results of a data artefact or true peaks. In this regard, it warrants mentioning that the data do become more volatile toward the end of the tracking period because fewer participants remain in the program.

In general, males and females did not appear to have large differences in terms of percentage of participants failing and average of fails per participant. Although both groups had a learning effect up to at least month ten in the program (when the data for females are more reliable), the descriptive analyses indicate that in general, the learning effect is more pronounced in male participants than female participants. The more volatile nature of the data for females is related to the smaller number of female participants in the program, especially after ten months. The logistic regression for the odds of failing tests versus passing tests supported these findings and even when the gender factor was found significant, the differences were very small.

Clear differences were found between mandatory and voluntary participants. Although both groups reveal a learning effect, the effect is more pronounced for voluntary participants. In general the mandatory participants exhibit odds for failing approximately 20% larger than the odds for the voluntary participants. Also, clear differences were found between participants with condition 37 and participants without this condition although both groups reveal a learning effect. The effect is more pronounced for the participants without the condition, who also have smaller odds for failing.

With respect to the comparisons between fails at start-up or fails running re-tests, it seems the learning effect is more pronounced at start-up with larger decreases in the percentage of participants failing, although it has to be acknowledged that the percentages of fails are larger at start-up than at running retests at the outset. Also, the odds for failing tests are larger at start-up than running re-retests.

With respect to the three BAC limits (0.02, 0.05 and 0.08), the odds for failing decrease faster for the more risky behaviours (fails over 0.05 and 0.08) suggesting that participants learn faster to be compliant for these more risky behaviours.

9. OVERALL CONCLUSIONS

The overall objective of the outcome evaluation was to examine the impact of Nova Scotia's interlock program on participants and to help identify areas for improvement. Different types of data were used in this evaluation, including conviction and crash records of individual participants, self-administered questionnaires to measure specific attitudes and behaviour, monthly counts of charges, convictions and crashes, and interlock logged events. Control data (not alcohol-related/non-interlocked participants) were also used to better support the findings for each type of experimental group (alcohol-related/interlock participants).

The data were analyzed in different ways. First, the descriptive analyses revealed that in general there were no significant differences between the respective experimental and control groups at the beginning of the study period with the exception of age-related differences. In terms of alcohol-related convictions, the control-voluntary group exhibited a recidivism rate of 8.9% during the study period, while the interlock-voluntary and interlock-mandatory groups had a 0.9% and 3% recidivism rate respectively after the installation of the interlock device. The recidivism rate for the interlock groups increased to 1.9% (voluntary group) and 3.7% (mandatory group) after the device was removed from the vehicle, but they were still smaller than the rate for the control-voluntary group. This means that interlock participants had lower recidivism rates during installation as well as post-removal compared to non-interlock participants.

At the beginning of the study the majority of the mandatory participants (73.9%) were reportedly changing their drinking behaviour or working to prevent a relapse (action and maintenance stage), compared to the remainder 26.1% that were considering a change but were doing little if anything about it (pre-contemplation and contemplation stages). The percentage of voluntary-interlock participants that were reportedly changing their drinking behaviour or working to prevent a relapse was 56.3%. These results showed that offenders in the Nova Scotia program appeared to display a greater positive attitude in the program relative to findings in other studies (Nochajski and Stasiewics 2006; Wieczorek and Nochajski 2005), demonstrating that the program may be more positive in general.

In terms of alcohol-related crashes the control-voluntary group had a crash rate of 1.2% during the study period, while the interlock-voluntary and interlock-mandatory groups had a 0.6% and 0.8% rate respectively (although the differences were not statistically significant). A variety of survival analysis techniques were used to provide more insight into the effectiveness of the program in terms of recidivism and crashes. The results supported the notion that the interlock program was associated with a positive impact on reducing the risk for alcohol-related convictions, and there appeared to be no difference in this respect between mandatory and voluntary participants. With respect to crashes, the survival analyses did not show any statistically significant difference between any of the studied groups.

To bolster the findings from the survival analyses, time series analysis techniques were also used to study monthly counts of charges, convictions and crashes including a before and after period (i.e., before and after the implementation of Nova Scotia's interlock program). The results suggested that there were no permanent effects associated with the implementation of the program in terms of the number of alcohol-related charges and convictions. There were significant, albeit temporary effects in the first and seventh month following the implementation. These effects included a 13.32% decrease in the number of alcohol-related charges and a 9.93% decrease in the number of alcohol-related convictions in the first and seventh month respectively following the implementation of the program. With respect to crashes, there were no significant effects associated with the implementation of the program in relation to the number of alcohol-related crashes with fatal and serious injuries at the 5% level of statistical significance. However, there was a permanent effect at the 10% level of significance that represented a small decrease – from a statistical point of view – in the number of alcohol-related crashes every month since June 2009 (tenth month after the beginning of the program). Note that this corresponded to a decrease of one fatal or serious alcohol-related crash in approximately 33 years. This is perhaps not unexpected as to date most studies have not yet been able to definitively demonstrate a positive impact on crashes due to the small sample sizes and small programs resulting in lack of sufficient data.

The amount of data from the questionnaires at exit and follow-up was insufficient to draw meaningful conclusions with respect to changes in attitudes and opinions regarding the interlock program, drinking behaviour, and drink driving behaviour. However, an interesting reported fact was that there was evidence showing that some interlock participants in the mandatory group drove a non-interlocked vehicle while in the program. This evidence should be considered in light of existing evidence about the alternative to interlocks, i.e., licence suspension, and which shows that many suspended drivers continue to drive. As such, the evidence from this study supports the notion that interlocked offenders driving non-interlocked vehicles rarely happens indeed.

Finally, data were analyzed from the interlock devices, without comparing this to a control group. Overall the results suggested that there were learning curves illustrating that offenders were more likely to violate program rules at the beginning of program participation, but over time these violations decreased as offenders learned about, or experienced the consequences of program violations and the nuances associated with the functioning of, and compliance with, devices. In general, the curves were steepest at the beginning of program participation until approximately month 10, indicating that the learning effect may decrease or stop after a period of time. In general, males and females did not appear to have clear differences in terms of the percentage of participants failing and average number of fails per participant, although both groups had a learning effect. The learning effect was more pronounced in male participants than female participants. Clear differences were found between mandatory and voluntary participants. Although both groups revealed a learning effect, the effect was more pronounced among voluntary participants as well. In addition, clear differences were found between participants with condition 37 and participants without the condition although both groups revealed a learning effect. The effect was more pronounced for participants without the condition.

In sum, with respect to specific deterrence (i.e., referring to preventing recidivism) among individuals in the program there was strong evidence to suggest that participation in the interlock program reduced the risk of alcohol-related charges among participants. With respect to general deterrence (i.e., referring to a preventive effect on the entire population of drivers in Nova Scotia) there was a temporary decrease in the number of alcohol-related charges and convictions in the first and seventh month respectively with respect to the implementation of the program. There was also some weaker evidence (at the 10 % level of statistical significance) that there was a permanent decrease in the number of alcohol-related crashes with fatal and serious injuries every month since the tenth month after the beginning of the program.

When considering all of the evidence combined, it can be argued that the implementation of the interlock program had a positive impact on road safety in Nova Scotia and that it reduced the level of drink driving recidivism in the province. There were also some promising indications to suggest a decrease in the number of alcohol-related road traffic crashes and fatalities due to the interlock program, although this finding has to be confirmed with more data. In sum, the evidence suggested the interlock program was better at preventing harm due to alcohol-impaired driving than the alternative of not using the interlock program.

10. RECOMMENDATIONS

10.1 Continue the use of the interlock program in Nova Scotia

When collectively considering the evidence, one main recommendation was clearly substantiated, i.e., to continue the use of the interlock program in the province. While the evidence regarding the positive impact of the program on crashes may have been weaker and needs further bolstering, in particular with respect to the general deterrent effect of the program on the entire population of Nova Scotia as a whole, nonetheless, the evidence convincingly showed that the recidivism rate of interlocked offenders was lower than that of non-interlocked offenders. This was true, not only when the device was installed, but this positive effect extended beyond the time when the device was removed, both for voluntary and mandatory interlocked offenders. This finding was consistent with many other studies that have evaluated the impact of interlock programs.

A comparison of the results from this evaluation with other evaluation studies showed that the reduction in recidivism rates in Nova Scotia (79%-90% reduction) appeared to be at the high end of the spectrum. More than 10 evaluations of interlock applications have reported reductions in recidivism ranging from 35-90% (Elder et al. 2011; Marques et al. 2010; Voas and Marques 2003; Vezina 2002; Tippetts and Voas 1997; Coben and Larkin 1999; Raub et al. 2001). While it was not possible to test this hypothesis due to the lack of an appropriate control group where no treatment services were provided at all, the strong evidence regarding the impact of the interlock program during the time when the device was installed as well as after removal may be related to the fact that Nova Scotia's program included the provision of treatment services in combination with the use of the interlock program. Additionally, interlock offenders appeared more ready for change relative to general research findings, which may also be explained by the treatment component of the interlock program (note that all offenders included in this study, both interlocked ones and non-interlocked ones received some basic form of treatment). It is known from the literature that, generally speaking, better results are obtained when the interlock is used in combination with some form of treatment, rather than using the interlock by itself (Zador 2011).

10.2 Consider the systematic use of a performance-based exit in the interlock program

Despite the fact that the evidence in this study showed that the program had an effect that extended beyond the removal of the device, it is acknowledged that the evidence also showed that any benefits of the program seemed to diminish once the device was removed. For this reason, in combination with evidence from the literature regarding the usefulness of a performance-based exit whereby an offender's time on the device is extended until he/she can demonstrate compliance with program rules, another recommendation is to consider the systematic use of this feature in Nova Scotia's interlock program. While such a performance-based exit is already used on an ad-hoc basis in the program, it is recommended to formulate specific program rules that would enable the systematic use of this program feature. As such, an optimal balance can be achieved between rehabilitation and public safety.

10.3 Consider further strengthening of monitoring in the interlock program

The pronounced effects with respect to reduced recidivism rates and learning curves underscored the quality of the program in its current form. However, there was also some evidence suggesting that the program could further benefit from stronger monitoring still (Zador et al. 2011; Vanlaar et al. 2013, Casanova-Powell et al. 2014). For example, there was evidence to suggest that in rare occasions an interlocked offender drove a non-interlocked vehicle. While this evidence should be weighed in the context of the performance of alternative measures where non-compliance is typically not so exceptional (see for example the levels of disregard of license suspension that are traditionally high), these findings suggested the need for monitoring mileage levels of interlocked offenders to detect early any indications of non-compliance, face-to-face meetings with interlocked offenders at servicing during their time on the interlock to establish a rapport with them also to detect early instances of non-compliance, as well as sufficiently high levels of traffic enforcement in the province to establish a general deterrent effect that could help reduce the likelihood of such instances of interlocked offenders driving non-interlocked vehicles.

10.4 Consider focusing on levels of risk in relation to non-compliance

The evidence from the interlock data analysis suggested that learning curves were more pronounced in relation to riskier behaviours, i.e., failed tests at higher BAC limits (0.05 and 0.08). While it can be argued that this was a positive finding in itself, it also illustrated the need to provide clear feedback and education to offenders about the dangers and problems associated with drinking and driving at lower alcohol levels. Essentially, offenders on the interlock should not be drinking at all, and there are several reasons for this. Therefore, it is important they learn equally fast about compliance with regard to lower limits such as the 0.02 limit.

This was also true in relation to the high-risk offenders (mandatory ones and those with condition 37) who were less amenable to learning to be compliant.

10.5 Consider the continued monitoring of crash data

In light of the fact that the time series analysis found a permanent effect that was borderline-significant, i.e., only significant at the level of 10% but not at the more rigorous 5% level, it is paramount to continue to monitor these trends. In this regard, it warrants mentioning that only crash data through to 2010, inclusive were available for this time series analysis. It is possible that this permanent effect may turn out to be significant at the 5% level after all, should more years of data be available to strengthen analyses. Given that the ultimate goal of any road safety measure should be to decrease the number of crashes and victims, it goes without saying that it is important to update these analyses accordingly when data become available.

REFERENCES

Alcohol Indicators Report 2011 Nova Scotia. Nova Scotia department of Health and Wellness, 2011.

Asbridge, Mark and Donald Langille. (2013). Nova Scotia Student Drug Use Survey 2012, Technical report. Addiction Services, Nova Scotia department of Health and Wellness, and Dalhousie University.

Baker, A. (1987). In Car Alcohol Breath Analyzer – A Pilot Study. Safety Education Centre, College Park, Maryland.

Becketti, S. Introduction to Time Series Using Stata. Statacorp Lp, 2013.

Cleves, M., Gould, W., Gutierrez, R. G. (2003) An Introduction to Survival Analysis Using Stata. Stata Press; Revised edition (October 15, 2003)

Coben, J.H., and Larkin, G.L. (1999). Effectiveness of Ignition Interlock Devices in Reducing Drunk Driving Recidivism. *American Journal of Preventative Medicine* 16(1S): 81-87.

Changing the culture of alcohol use in Nova Scotia: an alcohol strategy to prevent and reduce the burden of alcohol-related harm in Nova Scotia. (2007) Department of Health Promotion and Protection, Addiction Services, Province of Nova Scotia.

Elder, R.W., Voas, R., Beirness, D., Shults, R.A., Sleet, D.A., Nichols, J.L., & Compton, R. (2011). Effectiveness of ignition interlocks for preventing alcohol-impaired driving and alcohol-related crashes. *American Journal of Preventative Medicine*, 40(3), 362-376.

Hodgins, David, C. (2001) Stages of change assessments in alcohol problems: Agreement across self and clinician reports. Substance Abuse, 22(2), pp. 87-96.

Linden, A., Adams, J.L. (2011) Applying a propensity-score based weighting model to interrupted time series data: improving causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, 17:1231-1238.

Marques, P.R., Voas, R.B., Roth, R., & Tippetts, S.A. (2010). *Evaluation of the New Mexico Ignition Interlock Program*. Report No. DOT HS 811 410. Washington, D.C.: National Highway Traffic Safety Administration.

McCleary, R., & Hay, R. (1980) Applied time series analysis for the social sciences. Sage Publications.

Nochajski TH. Training Manual for the Research Institute on Addictions Self Inventory (RIASI)—Revised. Research Institute on Addictions; Buffalo, NY, USA, 2002.

Nochajski, T.H., & Stasiewicz, P.R. (2006). Relapse to driving under the influence (DUI): A review. *Clinical Psychology Review*, 26, 179-195.

Prochaska, J. O. and C. C. DiClemente (1986). Toward a comprehensive model of change. Addictive Behaviors: Processes of Change. W. R. Miller and N. Heather. New York, Plenum Press: 3-27.

Robertson, R., Holmes E., & Vanlaar, W. (2009). Alcohol Interlocks in Canada - From Research to Practice.

Robertson, R., Holmes E., & Vanlaar, W. (2010). Nova Scotia Alcohol Ignition Interlock Program: Process Evaluation.

Roberston, R., Vanlaar, W, Simpson, H., & Boase, P. (2009). Results from a national survey of Crown prosecutors and defense counsel on impaired driving in Canada: A "System Improvements" perspective. *Journal of Safety Research*, 40, pp. 25-31.

Royston, P. and M. K. B. Parmar. Flexible parametric models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in Medicine. 2001.

StataCorp. Stata: Release 13. Statistical Software. College Station, TX: StataCorp. LP, 2013.

Tippetts, A.S. and Voas, R.B. (1997). The effectiveness of the West Virginia interlock program on second drunk-driving offenders. In: C. Mercier-Guyon (Ed.) Alcohol, Drugs and Traffic Safety – T97. Proceedings of the 14th International Conference on Alcohol, Drugs and Traffic Safety, Annecy, France, September 21-26, 1997. Annecy: CERMT, Vol.1, pp. 185–192.

Vanlaar, W., McKiernan, A., Robertson, R. (2013). Behavioral Patterns of Interlocked Offenders: Phase II. Ottawa: Traffic Injury Research Foundation.

Vezina, L. (2002). The Quebec Alcohol Interlock Program: Impact on Recidivism and Crashes. In: D.R. Mayhew & C. Dussault (Eds.) Alcohol, Drugs and Traffic Safety 0 T2002. Proceedings of the 16th International Conference on Alcohol, Drugs and Traffic Safety. Montreal, August 4-9, 2002. Quebec City: Société de l'assurance automobile du Québec, pp. 97-104.

Voas, R.B. and Marques, P.R. (2003). Commentary: Barriers to Interlock Implementation. Traffic Injury Prevention 4(3): 183-187.

Wieczorek, W.F., & Nochajski, T. (2005). Characteristics of Persistent Drinking Drivers: Comparisons of First, Second, and Multiple Offenders. In D. Hennessey and D. Wiesenthal (Eds.), Contemporary Issues in Road User Behavior and Traffic Safety (pp. 153-166). New York: Nova Science Publishers Inc.

Yaffee, R. Introduction to time series analysis and forecasting, Academic Press. 200.

Zador, P. L., Ahlin, E. M., Rauch, W. J., Howard, J. M., & Duncan, G. D. (2011). The Effects of Closer Monitoring on Driver Compliance with Interlock Restrictions. Accid Anal Prev. Nov 2011; 43(6): 1960–1967

APPENDIX A: OFFENCE CODES

Alcohol related offence codes

Offence Description	Criminal Code Section
blood alcohol level over legal limit causing bodily harm	255(2.1)
blood alcohol level over legal limit causing death	255(3.1)
impaired causing bodily harm	255(2)
failure or refusal to provide sample causing bodily harm	255(2.2)
impaired causing death	255(3)
failure or refusal to provide sample causing death	255(3.2)
impaired operation	(253(1)(a) or 253(a)
blood alcohol over legal limit	253(1)(b) or 253(b)
failure or refusal to comply with demand	254(5)
Nova Scotia Motor Vehicle Act	
Offence Description	**Code**
newly licensed driver over zero BAC	100A
over legal limit or refusal	279A
low BAC	279C

Non-alcohol related offence codes

Criminal Code of Canada (CCC) Offences	
Offence Description	**Criminal Code Section**
Dangerous operation of motor vehicle (no injury)	249(1)(a)
Dangerous operation of motor vehicle (causing injury)	249(3)
Dangerous operation of motor vehicle (causing death)	249(4)
Driving while disqualified	259(4)
Failure to stop at the scene of an accident	252(1)(a)(b)(c)
Nova Scotia Motor Vehicle Act	
Offence Description	**Code**
Unlicensed driving	64, 80
Unsafe lane change and offences related to lined traffic	111,
Unsafe following distance	117(1)
Driving to left of centre line "Duty to drive on right"	110
Using hand-held cell phone or texting while driving	100D(1)
Seat belt violation	175(2)(3)(4)(5)(6)
Careless and imprudent driving	100
Speeding or dangerous driving	101
Passing school bus or failure to obey crossing guard	103(3), 125A(4)
Improper overtaking and passing	114,
Driving on left of centre line	115
Speeding in excess of speed limit (30 km/hr. and under)	102, 106A(a)(b), 106B(1)(a)(b)

Failure to obey traffic signs or signals or yield right of way	83(2)
Failure to yield to pedestrian	125(1)(2)
Manslaughter resulting from the operation of a motor vehicle in violation of section 236 CCC	278(1)(a)

APPENDIX B: QUESTIONNAIRES

- Demographics
- Self-reported behaviour
- Readiness to Change
- Research Institute on Addictions Self Inventory (RIASI)
- Expectations about Interlocks

Demographic Information

D1 Date of this interview

DD/MM/YYYY [ ]

D2 Driver licence number **(to be completed by clinician)**

[ ]

D3 Case ID number **(to be completed by clinician)**

[ ]

D4 First and Last name (please print)

[ ]

D5 What is your date of birth?

DD/MM/YYYY [ ]

D6 What is your marital status?

1 ☐ Single
2 ☐ Married
3 ☐ Living together
4 ☐ Divorced / separated
5 ☐ Widowed
6 ☐ Other status (write in) [ ]

D7 Who else lives in your household? **Select all that apply.**

1 ☐ Wife / partner
2 ☐ Children
3 ☐ Brothers / sisters
4 ☐ Friends
5 ☐ Parents / step-parents / other relative
6 ☐ Other
7 ☐ Live alone, **Go To question D10**

D8 Does anyone else in your household have a driving licence? **Select all that apply.**

1 ☐ Husband / wife / partner
2 ☐ Children
3 ☐ Brother / sister
4 ☐ Friend
5 ☐ Parent / step-parent / other relative
6 ☐ Other (write in) [ ]
7 ☐ No one else

D9 Is anyone else in your household with a driver's license restricted to using the interlocked vehicle?

1 ☐ Yes
2 ☐ No

Demographic Information

D10 How many motor vehicles do you have readily available for your personal use?

\# of vehicles [ ]

D11 Which of the following best describes your current position? **Select only one category.**

1 [ ] Self-employed
2 [ ] Employed part-time
3 [ ] Employed full-time
4 [ ] Housewife / husband
5 [ ] Retired
6 [ ] Unemployed and seeking work
7 [ ] Unemployed but not seeking work
A [ ] Student
B [ ] Other (write in) [ ]

D12 To which of these groups do you consider you belong?

1 [ ] White
2 [ ] Black-Caribbean
3 [ ] Black-African
4 [ ] Black-other black groups
5 [ ] Indian
6 [ ] Pakistani
7 [ ] Chinese
0 [ ] None of these

D13 Was the offence that led to your recent disqualification your first drink driving conviction?

1 [ ] Yes
2 [ ] No

D14 Do you feel you were given enough information about the interlock program in order to decide whether you would want to participate?
(Please circle the number you most agree with)

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

D15 Please check the box that you most identify with

1 [ ] I am/will be in the interlock program
2 [ ] I am/will **not** be in the interlock program

Self-Reported Behaviour-Intake

S1 On how many occasions in the last month have you needed to drive your car while you were drinking or within one hour of drinking but decided **not** to drive it?

1 ☐ Enter number of occasions

S2 On how many occasions in the last month have you needed to drive your car while you were drinking or within one hour of drinking and decided to drive it?

1 ☐ Enter number of occasions,
If response is zero (0), **please move on to the next questionnaire**

S3 The last time when you drove your car while you were drinking or within one hour of drinking, where were you doing most of your drinking?

1 ☐ Bar
2 ☐ Restaurant
3 ☐ Own home
4 ☐ Friend / relative's house
5 ☐ Party
6 ☐ Other (write in) ____________

Self-Reported Behaviour – Exit, follow-up Interlock

S1 On how many occasions in the last month have you needed to drive your car while you were drinking or within one hour of drinking but decided **not** to drive it?

1 ☐ Enter number of occasions

S2 On how many occasions in the last month have you needed to drive your car while you were drinking or within one hour of drinking and decided to drive it?

1 ☐ Enter number of occasions,

If response is zero (0), **Go to question S4**

S3 The last time when you drove your car while you were drinking or within one hour of drinking, where were you doing most of your drinking?

1 ☐ Bar
2 ☐ Restaurant
3 ☐ Own home
4 ☐ Friend/relative's house
5 ☐ Party
6 ☐ Other (write in) ________

S4 On how many occasions during your participation in the interlock program have you driven a non-interlock vehicle?

1 ☐ Enter number of occasions

S5 Do you think it is likely that you will drink and drive again in the future now that the interlock will no longer be installed in your car?

1 ☐ Yes
2 ☐ No

S6 Do you think you will plan ahead in the future to arrange for alternative transportation next time when you are drinking?

1 ☐ Yes
2 ☐ No

Self-Reported Behaviour –Exit DWI

S1 On how many occasions in the last month have you needed to drive your car while you were drinking or within one hour of drinking but decided **not** to drive it?

1 ☐ Enter number of occasions

S2 On how many occasions in the last month have you needed to drive your car while you were drinking or within one hour of drinking and decided to drive it?

1 ☐ Enter number of occasions

If response is zero (0), **Go to question S4**

S3 The last time when you drove your car while you were drinking or within one hour of drinking, where were you doing most of your drinking?

1 ☐ Bar
2 ☐ Restaurant
3 ☐ Own home
4 ☐ Friend / relative's house
5 ☐ Party
6 ☐ Other (write in) ____________

S4 Do you think it is likely that you will drink and drive again in the future now that your revocation is over?

1 ☐ Yes
2 ☐ No

S5 Do you think you will plan ahead in the future to arrange for alternative transportation next time when you are drinking?

1 ☐ Yes
2 ☐ No

Readiness to Change

This questionnaire asks for your thoughts about your own drinking. Please check to show how strongly you AGREE or DISAGREE with each statement.

		Strongly agree	Agree	Not sure	Disagree	Strongly disagree
RE1	I don't think I drink too much					
RE2	I am trying to drink less.					
RE3	I was drinking too much at one time but I have managed to change					
RE4	I enjoy drinking but I feel I drink too much					
RE5	I sometimes think I should cut down on drinking					
RE6	I have changed my drinking but I am looking for ways to keep from slipping back to the old pattern					
RE7	I feel that it is a waste of time talking about drinking					
RE8	I have recently changed my drinking					
RE9	I want to keep from going back to the drinking problem I had before					
RE10	I am actually doing something about my drinking					
RE11	I feel I should consider drinking less					
RE12	I feel that drinking is a problem sometimes					
RE13	I feel that there is no need for me to change my drinking					
RE14	I am changing my drinking habits					
RE15	I feel it would be pointless to drink less					
RE16	I see myself as an alcoholic					

RIASI

Please check the box to show if the statement is true or false

		True	False
RI1	I smoke or use tobacco products		
RI2	I have no problem telling a companion that he or she has done something to hurt my feelings		
RI3	I often feel so restless I can't sit still		
RI4	When I drank 7 or more drinks I become aggressive		
RI5	I like people who are sharp and witty even though they may sometimes hurt other peoples` feelings		
RI6	When the alcohol runs out, I leave a party		
RI7	When I make plans, I am almost certain to make them work		
RI8	I have relatives who have had problems with alcohol or drugs		
RI9	I have been arrested for crimes other than drinking and driving		
RI10	My hand often shakes when I try to do something		
RI11	I am irritated a great deal more than people are aware of		
RI12	Since the age of 18, I have been accidentally cut, or cut in a fight, or burned badly enough to leave a scar		
RI13	A family member was arrested for drinking and driving		
RI14	When I don't got my own way, I sulk or pout		
RI15	I slow down when a traffic light turns to amber		
RI16	I often feel like a powder keg ready to explode		
RI17	When I have a problem I try to make it go away by drinking		
RI18	I have no trouble sleeping or staying asleep		
RI19	I sometimes do dangerous or risky things just for fun		
RI20	I have experienced a major stressful life event in the past 12 months		
RI21	I feel that I have lived the right kind of life		
RI22	It Is easy for me to turn down an unreasonable request from a friend		
RI23	I have feelings that something bad will happen to me		
RI24	I feel like I have lost energy I am fatigued and tired		
RI25	I often have feelings of nervousness		
RI26	I often feel sad or blue		
RI27	A drink or two gives me energy to get started		
RI28	I am probably not capable of slapping someone, even when I lose my temper		
RI29	When I get beyond a certain point, I don't stop drinking until all the booze is gone or I pass out		
RI30	I don't like to break rules, even if I think they are wrong		

		True	False
RI31	I hardly ever drink more than I plan to		
RI32	I am not interested in surprising or upsetting others by doing something that might shock them		
RI33	It depresses me that I did not do more for my parents		
RI34	I like to gamble for money		
RI35	After seven or more drinks, I feel happier		
RI36	I often acted without thinking as a child		
RI37	I was referred for a liver test, or a blood test for liver enzymes		
RI38	Since the age of 18, I have needed emergency treatment for an injury of some kind		
RI39	I skipped school as a child		
RI40	When I am drinking, I make sure I do not skip any meals		
RI41	I often feel hopeless about the future		

Please answer the following question by writing in your response in the box next to each question.

RI42	How many jobs have you had in the past five years?	
RI43	How many times have you ever been convicted for moving traffic offences such as speeding, running a red light or failing to stop at a STOP sign?	
RI44	How much money do you usually spend on alcohol in a week? (Include the cost of drinking at home, at friends' or relatives' houses and in pubs, bars and restaurants)	
RI45	If you go out drinking, how many places do you drink at in one evening? (Include friends' and relatives' houses as well as pubs, bars and restaurants)	
RI46	What is the largest number of drinks you have ever consumed in a 24 hour period? (One drink is a ½ pint beer/lager, a single measure of spirits, a glass of wine or one Alcopop.)	
RI47	How many days of the week do you usually drink?	
RI48	When you are drinking, how many drinks do you usually have?	
RI49	How many drinks does it take before you begin to feel the effects of alcohol?	

Listed below are a few statements about your relationships with others.
Please circle the number to indicate how much each statement is TRUE or FALSE for you.

		Definitely true	Mostly true	Don't know	Mostly false	Definitely false
RI50	I am always courteous even to people who are disagreeable	1	2	3	4	5
RI51	I sometimes feel resentful when I don't get my way	1	2	3	4	5
RI52	No matter who I'm talking to, I'm always a good listener	1	2	3	4	5

Expectations about Interlocks-Intake

This questionnaire asks about your expectations of the interlock. Please check to show how strongly you AGREE or DISAGREE with each statement.

		Strongly agree	Agree	Not sure	Disagree	Strongly disagree
E1	I am sure that I will be able to use the Interlock OK					
E2	The Interlock will stop me from driving after drinking.					
E3	Using the Interlock will be embarrassing for me.					
E4	Having the Interlock fitted in the car will be embarrassing for me					
E5	I think using the Interlock will become a habit, just like putting your seat belt on					
E6	Having the Interlock will allow me to keep my job.					
E7	Having the Interlock will keep me from becoming dependent on others for transport.					
E8	Having the Interlock will help maintain family harmony.					
E9	Using the Interlock could become a hassle.					
E10	Having the Interlock will affect my drinking habits.					
E11	Having the Interlock will change my driving habits.					
E12	I expect to benefit from using the Interlock.					
E13	I think the advantages of using the Interlock are greater than the disadvantages.					
E14	I wouldn't bother to try and beat the Interlock – it's got too many security measures					

Expectations about Interlocks – Exit, follow-up

This questionnaire asks about your previous experiences with the interlock. Please check to show how strongly you AGREE or DISAGREE with each statement.

		Strongly agree	Agree	Not sure	Disagree	Strongly disagree
E1	I was able to use the Interlock OK					
E2	The Interlock stopped me from driving after drinking					
E3	Using the Interlock was embarrassing for me					
E4	Having the Interlock fitted in the car was embarrassing for me					
E5	I think using the Interlock became a habit, just like putting your seat belt on					
E6	Having the Interlock allowed me to keep my job					
E7	Having the Interlock kept me from becoming dependent on others for transport					
E8	Having the Interlock helped maintain family harmony					
E9	Using the Interlock became a hassle					
E10	Having the Interlock affected my drinking habits					
E11	Having the Interlock has changed my driving habits					
E12	I benefitted from using the Interlock					
E13	I think the advantages of using the Interlock are greater than the disadvantages					
E14	I wouldn't bother to try and beat the Interlock – it's got too many security measures					