MODELING THE BEHAVIOR OF NOVICE YOUNG DRIVERS USING DATA FROM IN-VEHICLE DATA RECORDERS

Tsippy Lotan[1], Tomer Toledo[2], Carlo G. Prato[3]
[1]Or Yarok, Hod Hasharon, Israel
[2]Faculty of Civil and Environmental Engineering
Technion - Israel Institute of Technology, Haifa, Israel
[3]Ran Naor Road Safety Research Center
Technion - Israel Institute of Technology, Haifa, Israel
Email: tsippy@oryarok.org.il; toledo@technion.ac.il; prato@technion.ac.il

Summary: Novice young drivers suffer from increased crash risk that translates into over-representation in road injuries. A better understanding of the driving behavior of novice young drivers and of their determinants is needed to tackle this problem. To this extent, this study analyzes the behavior of novice young drivers within a Graduated Driver Licensing (GDL) program. Data on driving behavior of novice drivers and their parents is collected using in-vehicle data recorders, which calculate compound risk indices as measures of the risk taking behavior of the various drivers. Data is used to estimate a negative binomial model to identify the major factors that affect the driving behavior of the young drivers. Estimation results suggest that the risk taking behavior of young drivers is influenced by that of their parents and decreases with higher levels of supervised driving and stricter monitoring by the parents.

OBJECTIVES

Novice young drivers suffer from increased crash risk that translates into over-representation in road injuries. Many jurisdictions worldwide have introduced Graduated Driver Licensing (GDL) programs to tackle this problem. These programs are designed to allow young drivers to obtain as much practical driving experience in real-world conditions as possible and, at the same time, to limit their exposure to the high-risk situations inherent to the novice driver status.

In recent years, several studies analyzed the effectiveness of GDL programs in reducing novice drivers' crash rates at the single (e.g., Foss, Feaganes and Rodgman, 2001; Shope and Molnar, 2004) and multi-jurisdictional level (e.g., Dee, Grabowski and Morrisey, 2005; Chen, Baker and Li, 2006). Other studies focused on specific aspects such as the importance of parental involvement in teen driver licensing (e.g., Simons-Morton et al., 2006; Prato, Lotan and Toledo, forthcoming), the relevance of seat-belt use (Goodwin and Foss, 2004; Goodwin et al., 2006), and the necessity of passenger and night-time restrictions (e.g., Simons-Morton, Lerner and Singer, 2005; Morrisey et al., 2006). Existing research provides evidence that the implementation of GDL policies establishes a real-world learning environment, which is relatively safe. However, it is not clear whatever occurs during the initial months of driving to bring young drivers' crash rates down fairly dramatically (Foss, 2007).

This study analyzes the behavior of novice young drivers within a GDL program by using data collected with in-vehicle data recorders (IVDR). The IVDR, which were installed in vehicles

driven by novice young drivers, measure and elaborate speed and acceleration data to identify various driving maneuvers that the vehicle undertakes and their severity. These maneuvers provide unbiased and objective observations of driving behavior. The driving maneuvers recorded for each driver are combined to calculate risk indices for the novice drivers and their family members who drive the instrumented vehicle. These risk indices are used as indicators to the drivers' risk taking behavior in estimating a model that aims to capture the factors that affect this behavior.

METHODS

Data collection

Data were collected with an IVDR system developed by GreenRoadTechnology. The IVDR was installed in the vehicles that participated in the study. For all trips made with the vehicle, the IVDR identifies the driver, registers trip start and end times, and records speed and acceleration measurements at high time resolution. Pattern recognition algorithms are used to identify various maneuvers (e.g., lane changes, turns with and without acceleration, sudden brakes, strong accelerations, high speeds) in the raw speed and acceleration profiles. The maneuvers are also classified in three levels of severity (moderate, intermediate or risky) based on characteristics such as their duration and extent of sudden changes in speed and acceleration. The processed information is sent through wireless networks to an application server, which maintains a database with vehicle-specific and driver-specific trip history consisting of statistics of the vehicle usage patterns, recorded maneuvers and severity ratings. A complete description of this IVDR system is presented by Toledo, Musicant and Lotan (2008).

According to the Israeli GDL program implemented in 2000, teenagers are allowed to start taking on-road driving lessons with professional instructors at the age of 16.5 years. Learners are not allowed to drive outside these lessons. Before taking the on-road driving test, the learners must pass a theoretical test, be at least 17 years old and have attended at least 28 driving lessons. For the first three months after licensure, named the accompanied driving period, the novice drivers must be accompanied whenever they drive by an experience driver, who is at least 24 years old and holds a valid driving license for at least five years. In the following period and up to two years after licensure, the novice drivers are subject to passenger limitations unless an experienced driver is present in the vehicle. Further details on the Israeli GDL program are provided by Lotan and Toledo (2007).

Participants in this study were selected from volunteer families of newly licensed drivers. The families were recruited through advertisements in a dedicated web-site, in the media and in professional driving schools. Initially, families were screened to verify that most or all the trips of the newly licensed driver would be on the vehicle where the IVDR was installed, and that this vehicle was the main vehicle used by the accompanying persons. Collection of driving behavior history of young drivers and their family members took place over a period of twelve months that included both the initial three months of the accompanied driving period, and the nine months of the solo driving period that follows. In order to minimize the effect of the system on their behavior, participating families received initially only minimal information about the purpose and capabilities of the IVDR and no feedback at all about the observed driving behavior.

Approximately four months after the IVDR installation, the participants received additional information on the IVDR and access codes to personal web pages containing the data collected on their driving behavior and risk indices.

Data analysis and model estimation

Recorded maneuvers and their severity levels were aggregated to calculate risk indices for the novice driver and for the family members driving the same vehicle on a monthly basis. These indices are proxies for the driver's risk of involvement in car crashes and have been shown to be positively correlated with drivers' actual crash records (Toledo, Musicant and Lotan, 2008). Risk indices are expressed as a linear function of the number and severity of the maneuvers in each month, normalized by the driving time in that month:

$$R_{im} = \frac{E_{im}}{DT_{im}} = \frac{\sum_j \sum_s \beta_{js} N_{ijsm}}{DT_{im}} \qquad (1)$$

where R_{im} is the risk index for individual i during month m and E_{im} is the equivalent number of events, which is calculated as a weighted sum of the number of maneuvers for the individual. DT_{im} is the driving time for individual i during month m, N_{ijsm} is the number of maneuvers of type j and severity level s for individual i during month m, and β_{js} are weights of the maneuvers of type j and severity level s.

The sample considered in this study consists of 68 families in which novice drivers drove the equipped vehicle for at least five hours per month during the experiment, and each parent drove the same vehicle for at least twenty-five hours overall. During the twelve-month period, 40 male and 28 female young drivers were monitored for a total of almost 10,000 driving hours in which they recorded nearly 49,000 maneuvers with intermediate or risky severity ratings. About 11,000 driving hours and 32,000 maneuver events were recorded for the 68 fathers and mothers of the novice drivers. Note that, since participation in the study was on a voluntary basis, the sample is not representative of the Israeli population and is likely to be biased towards self-selection of families with high awareness and willingness to participate. Data aggregation and risk index calculation on monthly basis yielded an unbalanced panel of 573 observations for the 68 young drivers, given that not all novices drove during all the months of the experiment.

The IVDR data were used to develop a model that explains the risk taking behavior of the novice drivers, as captured by the risk indices during the first year after licensure, and the factors that affect it. The dependent variable in the model is the monthly weighted number of events E_{im}, which is available for a panel of N young drivers over M periods. Thus, count data models for panels were estimated. Hausman specification tests established insignificant correlations between unobserved person-specific random effects and regressors. Therefore, the random effects model was more powerful and parsimonious than the fixed effects model. Following Hausman, Hall and Griliches (1984), conditional on the driver-specific random effects the model is initially specified as a traditional Poisson regression model:

$$\log E_{im} = \log DT_{im} + \beta' X_{im} + u_i + \varepsilon_{im} \qquad (2)$$

where *log DT$_{im}$* is the exposure variable that normalizes the events by the driving time of young driver *i* during month *m*. *X$_{im}$* and *β* are a matrix of explanatory variables and the corresponding vector of parameters to be estimated, respectively. *u$_i$* is a random effect for driver *i* and *ε$_{im}$* is a Gamma distributed error term.

The inclusion of the driver-specific random terms *u$_i$* creates over-dispersion, which is randomly distributed across drivers. Assuming that *u$_i$* are Gamma distributed in the population with parameters *(θ$_i$,θ$_i$)*, the unconditional model is a negative binomial regression model with an over-dispersion parameter that varies across drivers. The ratio *θ$_i$/(1+θ$_i$)* follows a beta distribution with parameters *(a,b)*.

The normalization of the events by the driving time allows rewriting the model in equation 2 as follows:

$$\log E_{im} - \log DT_{im} = \beta' X_{im} + u_i + \varepsilon_{im} \tag{3}$$

$$\log\left(\frac{E_{im}}{DT_{im}}\right) = \log R_{im} = \beta' X_{im} + u_i + \varepsilon_{im} \tag{4}$$

Maximum likelihood was used to estimate the parameters of the random-effect negative binomial model described above.
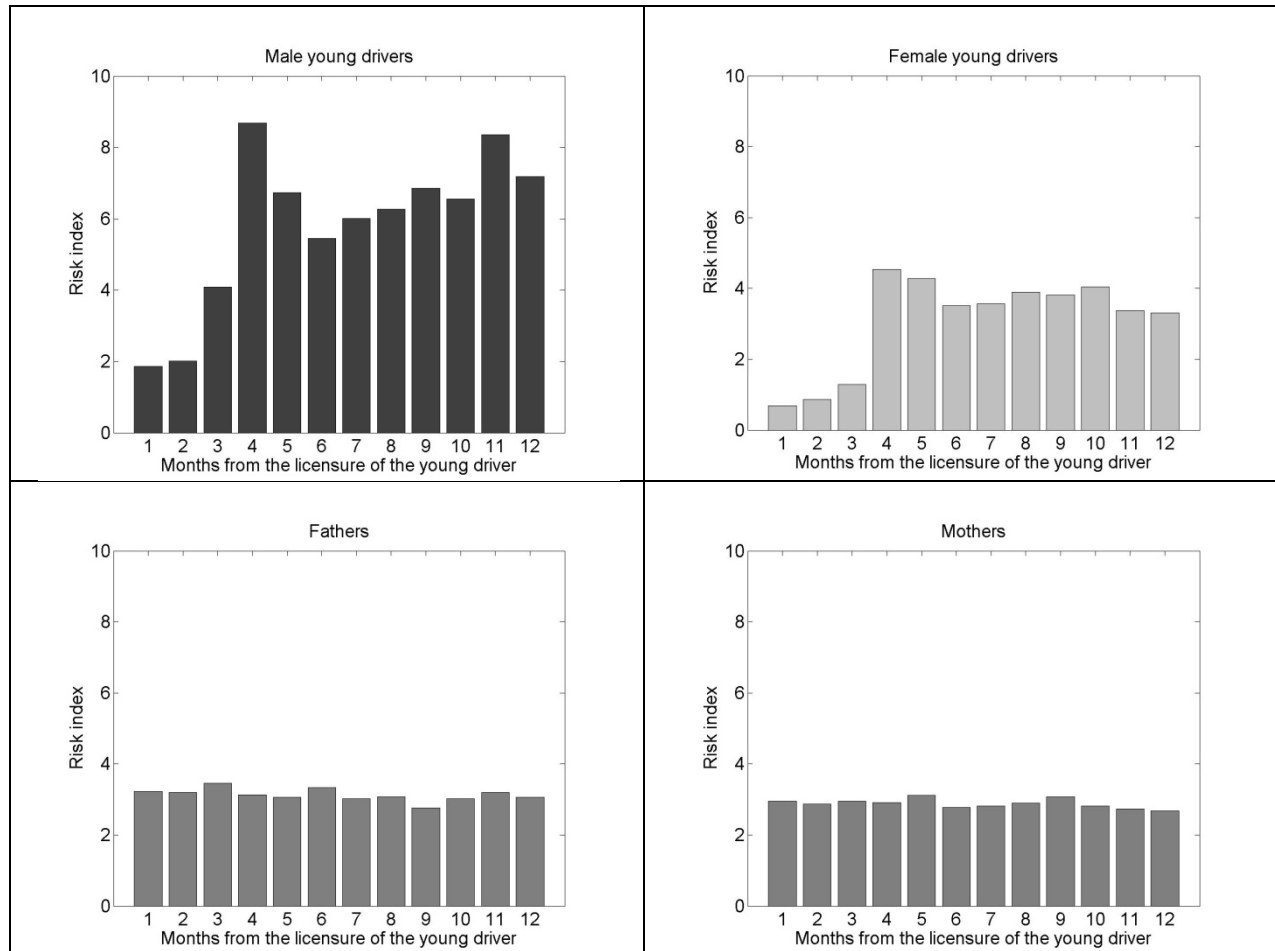
RESULTS

Risk indices

The variation over time of the average risk indices for novice drivers and their parents is presented in figure 1. For both male and female young drivers, risk indices vary significantly over time. Relatively low values are observed in the accompanied driving period. Steep increases are observed in the transition to the solo driving period. Feedback is provided to the families for the first time around the fourth or fifth month. At that time the average risk indices decrease substantially. This decrease is sustained over time for female drivers, but risk indices increase again for males. Overall, male novice drivers exhibit higher average risk indices compared to females. This result is in accordance with the existing literature that reports higher risk propensity in male drivers. Both male and female novice drivers score higher in the risk indices compared to their parents in the solo driving period, which is consistent with the higher crash rates observed for these drivers. Interestingly, the average risk indices of the parents do not change substantially during the twelve-month period, which may indicate that the driving styles of the parents are well established and stable.

Model estimates

Table 1 provides the definitions of the explanatory variables that were found statistically significant in the negative binomial model. Estimates of the parameters of the negative binomial model with random effects (eq. 2 and 4) are presented in Table 2.

Figure 1. Distribution of risk indices over time

The estimation results show that, as expected, males have higher risk indices. This is also consistent with the results in Figure 1. Risk-prone behavior of the parents, which is exhibited in their risk indices, is also associated with higher risk indices for the young drivers. This result is also consistent with previous findings (e.g. Taubman Ben-Ari, Mikulincer and Gillath, 2005; Prato, Lotan and Toledo, forthcoming).

In terms of the dynamic evolution of the risk indices, acquiring more hours of supervised driving hours reduced the risk index. In contrast, young drivers that drove more hours in the solo period (i.e. gained more driving experience) had higher risk indices compared to those that drove less. This result seems to indicate the importance of supervised driving experience as opposed to unsupervised driving to reduce risk-prone driving behavior.

The access to the feedback provided by the system also affects the risk indices of the young drivers. The risk indices decrease when the parents monitor the driving behavior of the young drivers. In contrast, risk indices increase if the young driver themselves (but not the parents) login to their driving records. A possible explanation may be that parents that use the feedback on the driving of their young drivers tend to monitor them more carefully and impose restrictions on their behavior.

Table 1. Explanatory variables in the negative binomial model

Variable	Values
$male_i$	1 if young driver i is male, 0 otherwise
$lnriskfather_i$	log of average risk index of the father of young driver i
$lnriskmother_i$	log of average risk index of the mother of young driver i
$loginyd_{im}$	1 if young driver i logs-in to the feedback in month m, 0 otherwise
$loginfm_{im}$	1 if the parents of young driver i log-in to the feedback in month m, 0 otherwise
$lndrivingtimeacc_{im}$	log of cumulative accompanied driving time in month m for young driver i
$lndrivingtimesolo_{im}$	log of cumulative solo period driving time in month m for young driver i

Table 2. Negative binomial model with random effects

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MALE	0.5737	0.0749	7.66	0.0000
LNRISKFATHER	0.3658	0.0427	8.57	0.0000
LNRISKMOTHER	0.3656	0.0575	6.36	0.0000
LOGINYD	0.1451	0.1010	1.44	0.1507
LOGINFM	-0.1769	0.0887	-1.99	0.0462
LNDRIVINGTIMEACC	-0.1162	0.0326	-3.57	0.0004
LNDRIVINGTIMESOLO	0.1859	0.0224	8.31	0.0000
CONSTANT	-0.2832	0.1129	-2.51	0.0121
EXPOSURE	1.0000	fixed parameter		
Beta distribution of the dispersion parameter for count data model				
a	1.8023	0.4502	4.00	0.0001
b	18.5132	4.5314	4.09	0.0000
Log-likelihood	-2516.31	Chi Squared		20030.15
Restricted (Poisson) Log likelihood	-12531.39	Degrees of freedom		1
Restricted (Constants) Log likelihood	-57583.67	Prob(Chi Squared>value)		0

CONCLUSIONS

This paper analyzed the behavior of novice young drivers in the first year after licensure. The data was collected using an IVDR system that continuously monitors the vehicle. These measurements, which provide objective observations of driving behavior, are used to calculate risk indices that have been previously shown to be positively correlated with the risk of crash involvement. A random effect negative binomial regression model was estimated to explain the monthly risk indices. The results help identify and provide insights on the important factors that affect these risk indices.

First, the results stress the role of parents in influencing the driving behavior of their children. This is evident both in the connection between the risk indices of parents and their children, and through the impact of parental monitoring to reduce the risk indices of their children. Thus, parents should be encouraged to provide positive driving behavior modeling and to actively monitor their young drivers' behavior and tailor family policies to discourage risky behaviors that potentially lead to car crashes. Furthermore, parents also have an important role in increasing the amount of supervised driving the novice drivers undertake, which seems to reduce subsequent risky behavior. This result may also imply that policies to extend the accompanied driving period and encourage accompanied driving (e.g. through setting minimum supervised driving hours requirements) could be useful to reduce crash risks of young drivers.

REFERENCES

Chen, L.H., Baker, S.P., & Li, G. (2006). Graduated driver licensing programs and fatal crashes of 16-year-old drivers: a national evaluation. *Pediatrics*, *118*(1), 56-62.

Dee, T.S., Grabowski, D.C., & Morrisey, M.A. (2005). Graduated driver licensing and teen traffic fatalities. *Journal of Health Economics*, *24*(3), 571-589.

Foss, R.D. (2007). Improving graduated driver licensing systems: a conceptual approach and its implications. *Journal of Safety Research*, *38*(2), 185-192.

Foss, R.D., Feaganes, J.R., & Rodgman, E.A. (2001). Initial effects of graduated driver licensing on 16-year-old driver crashes in North Carolina. *Journal of the American Medical Association*, *286*(13), 1588-1592.

Goodwin, A.H., & Foss, R.D. (2004). Graduated driver licensing restrictions: awareness, compliance, and enforcement in North Carolina. *Journal of Safety Research*, *35*(4), 367-374.

Goodwin, A.H., Wells, J.K., Foss, R.D., & Williams, A.F. (2006). Encouraging compliance with graduated driver licensing restrictions. *Journal of Safety Research*, *37*(4), 343-351.

Hausman, J., Hall, B.H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents R&D relationship. *Econometrica*, *52*(4), 909-938.

Lotan, T., & Toledo, T. (2007). Driving patterns of young drivers within a Graduated Driver Licensing system. Presented at the 86th Annual Meeting of the Transportation Research Board, Washington, D.C.

Morrisey, M.A., Grabowski, D.C., Dee, T.S., & Campbell, C. (2006). The strength of graduated drivers license programs and fatalities among teen drivers and passengers. *Accident Analysis and Prevention*, *38*(1), 135-141.

Prato, C.G., Lotan, T., & Toledo, T., (forthcoming). Intra-familial transmission of driving behavior: evidence from in-vehicle data recorders. *Transportation Research Record* (in press).

Shope, J.T., & Molnar, L.J. (2004). Michigan's graduated driver licensing program: evaluation of the first four years. *Journal of Safety Research*, *35*(3), 337-344.

Simons-Morton, B.G., Hartos, J.L., Leaf, W.A., & Preusser, D.F. (2006). Increasing parent limits on novice young drivers: cognitive mediation of the effect of persuasive messages. *Journal of Adolescent Research*, *21*(1), 83-105.

Simons-Morton, B.G., Lerner, N., & Singer, J. (2005). The observed effects of teenage passengers on the risky driving behavior of teenage drivers. *Accident Analysis and Prevention*, *37*(6), 973-982.

Toledo, T., Musicant, O., & Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behavior. *Transportation Research Part C*, *16*(3), 320-331.