Brief Original Contribution

Do Elevated Gravitational-Force Events While Driving Predict Crashes and Near Crashes?

Bruce G. Simons-Morton*, Zhiwei Zhang, John C. Jackson, and Paul S. Albert

* Correspondence to Dr. Bruce G. Simons-Morton, Prevention Research Branch, Division of Epidemiology, Statistics, and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 7B13M, Bethesda, MD 20892-7510 (e-mail: mortonb@mail.nih.gov).

The purpose of this research was to determine the extent to which elevated gravitational-force event rates predict crashes and near crashes. Accelerometers, global positioning systems, cameras, and other technology were installed in vehicles driven by 42 newly licensed Virginia teenage drivers for a period of 18 months between 2006 and 2009. Elevated gravitational force and crash and near-crash events were identified, and rates per miles driven were calculated. (One mile = 1.6 km.) The correlation between crashes and near crashes and elevated gravitational-force event rates was 0.60. Analyses were done by using generalized estimating equations with logistic regression. Higher elevated gravitational-force event rates in the past month substantially increased the risk of a crash in the subsequent month (odds ratio = 1.07, 95% confidence interval: 1.02, 1.12). Although the difference in this relation did not vary significantly by time, it was highest in the first 6 months compared with the second and third 6-month periods. With a receiver operating characteristic curve, the risk models showed relatively high predictive accuracy with an area under the curve of 0.76. The authors conclude that elevated gravitational-force event rates can be used to assess risk and to show high predictive accuracy of a near-future crash.

accident prevention; adolescent behavior; area under the curve; motor vehicle crashes; prediction; receiver operating characteristic; risk-taking; safety

Abbreviations: AUC, area under the curve; CNC, crash and near crash; ROC, receiver operating characteristic.

Elevated gravitational-force events due to changes in acceleration from late braking, rapid starts, and sharp turns are dangerous because they can increase the potential for loss of vehicle control and reduce the time available for drivers to respond to hazards and for other road users to respond to the drivers' behavior (1, 2). Wahlberg (3) has argued that the number of crashes is roughly equal to the sum of drivers' speed changes during the period preceding the crashes. Although a variety of studies have documented associations between crash risk and elevated gravitational-force event rates and other risky driving behaviors, the extent to which elevated gravitational-force events predict crashes has not been demonstrated.

The feasibility of studying the relation between driving behavior and crash risk has advanced since the advent of in-vehicle data recording devices equipped with accelerometers enabled objective measurement of these variables. Bagdadi and Varhelyi (1) reported a significant association between acceleration events and self-reported crash experience. Wahlberg (3) demonstrated that elevated gravitational-force events were associated with at-fault crashes among Swedish bus drivers. Klauer et al. (4) reported that drivers with higher crash and near-crash (CNC) rates also had higher rates of elevated gravitational-force events compared with the lower CNC group.

The driving behavior of teenage drivers is of particular interest because crash rates are high for younger drivers, particularly early in licensure (5–7). Speeding, acceleration, and other aggressive and risky driving behavior are commonly identified as contributing factors to crashes among young

Table 1. Gravitational-Force Event Prevalence and Correlation With CNCs, Naturalistic Teenage Driving Study, 2006–2009

Category	Gravitational Force	Frequency	% of Total Events	Alpha Coefficient	Correlation With CNCs
Rapid starts	>0.35	8,747	39.6	0.86	0.28
Hard stops	≤−0.45	4,228	19.1	0.59	0.76
Hard left turns	≤−0.05	4,563	20.6	0.83	0.53
Hard right turns	≥0.05	3,185	14.4	0.63	0.62
Yaw[a]	6 degrees in 3 seconds	1,367	6.2	0.57	0.46
Total (composite)		22,090	100	0.78	0.60

Abbreviation: CNCs, crashes and near crashes.

[a] "Yaw" is a measure of correction after a turn and is calculated as the delta-V (change in velocity) between an initial turn and the correction or swerve.

drivers (5). The Naturalistic Teenage Driving Study examined the driving behavior of newly licensed teenage drivers (8). The purpose of the research reported in this paper is to examine the prediction of CNCs by elevated gravitational-force event rates among novice young drivers. Specific research questions included the following: 1) Are elevated gravitational-force rates and CNC rates associated? 2) Do elevated gravitational-force rates predict the likelihood of a crash? 3) Does the relation vary over time?

MATERIALS AND METHODS

Study population

A convenience sample of newly licensed Virginia teenagers (22 females and 20 males; average age, 16.4 years) and at least 1 of their parents was recruited. Teenage participants received an incentive of $75 per month. Parent participants provided written informed consent, and teenage participants provided written assent. The research was reviewed and approved by the Institutional Review Board of the Virginia Technical Institute.

Data collection and measures

Vehicle instrumentation. Within 2 weeks of teen licensure, the participants' own vehicles were equipped with instruments that included a data recording system comprising a computer that received and stored continuous data from accelerometers to assess kinematic data, a global positioning system to assess mileage, and video recorders. Video cameras were installed on the rear view mirror aimed at the driver and forward and rearward roadway. Data for each teenage driver were collected for 18 months from June 2006 to September 2009. Data were downloaded periodically from the computers installed in the vehicle trunks. Video data for each vehicle trip were viewed by coders, and the identities of the occupants were coded.

Crashes and near crashes. Coders viewed video footage of highly elevated gravitational-force events and classified each CNC as driver at fault or not at fault. Because crashes are relatively rare, analyses were performed on rates of at-fault CNCs. Guo et al. (9) demonstrated that near crashes can serve as valid and useful surrogates for crashes.

Elevated gravitational-force event rates. Five categories of events were counted at specific gravitational levels, and a composite measure of the 5 events was created (Cronbach's alpha of 0.78) (Table 1).

Statistical methods

The distributions of individual at-fault CNCs and gravitational-force event rates were examined graphically. Spearman's correlation coefficients between CNC rates and gravitational-force event rates were calculated.

A logistic regression model was used to associate the occurrence of at-fault CNCs in a month with gravitational-force event rates immediately prior to the month as well as other relevant information. Generalized estimating equations (10) were used to account for the within-subject correlation among different months (excluding CNC data for the first month of each subject), with the logarithm of the number of miles driven in the same month as an offset. (One mile = 1.6 km.) The final model was determined after extensive exploratory analyses that included different forms of gravitational-force event rates (as 5 separate rates or 1 composite measure), time since licensure (in half-years), and prior CNC rates.

The predictive ability of the final logistic regression model was assessed by using receiver operating characteristic (ROC) curves (11). An ROC curve is a plot of sensitivity versus $1 -$ specificity corresponding to different cutoff values for a continuous predictor. The ROC curve for a useless predictor based on random guess and no relevant information is the diagonal in the unit square going from (0, 0) to (1, 1). The ROC curve for a perfect predictor consists of the left and upper edges of the unit square, which represent the highest possible path from (0, 0) to (1, 1). A succinct summary of an ROC curve is the area under the curve (AUC), which is 0.5 for a useless predictor and 1.0 for a perfect predictor. To avoid overfitting, we adopted a cross-validation approach comparing the outcome for each driver-month with the prediction based on the rest of the data (12). A bootstrap procedure was used to measure the sampling variability of the cross-validated AUC.
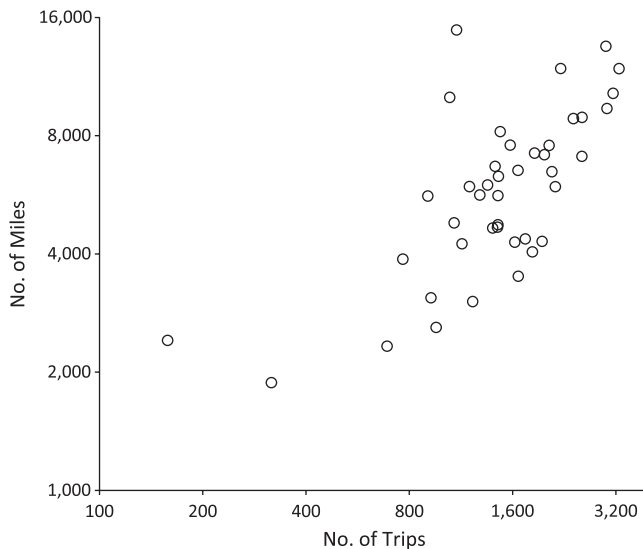
Figure 1. Summary of data by subject: the number of trips made and the total number of miles driven, both plotted on the logarithmic scale, Naturalistic Teenage Driving Study, Blacksburg, Virginia, 2006–2009. One mile = 1.6 km.
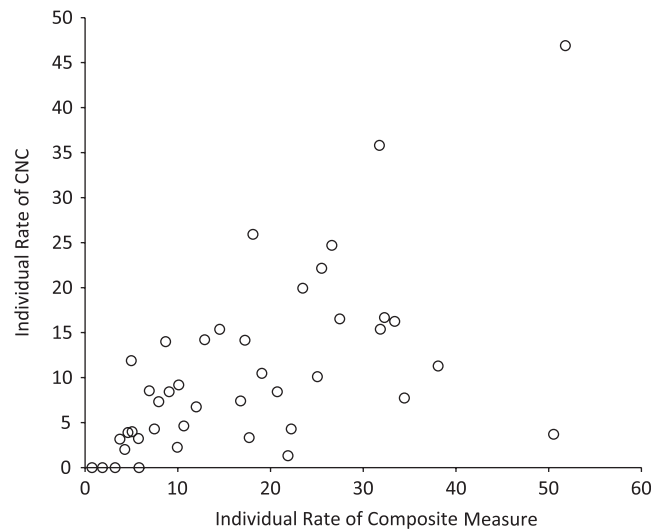

Figure 2. Individual event rates for crashes and near crashes (CNCs) (per 10,000 miles) and the composite measure of gravitational-force events (per 100 miles) (Spearman's $r = 0.60$), Naturalistic Teenage Driving Study, Blacksburg, Virginia, 2006–2009. One mile = 1.6 km.

RESULTS

The Naturalistic Teenage Driving Study data set comprises more than 68,000 trips with an average of 1,626 trips per subject. Figure 1 provides a summary by subject of the number of trips made and the total number of miles driven. A total of 279 CNC events occurred during those trips, of which 37 were crashes and 173 were at fault.

Correlation of CNCs with elevated gravitational-force event rates

Figure 2 presents a scatter plot of individual event rates for CNCs and the composite measure of gravitational-force event rates, showing skewed distributions. The composite measure of elevated gravitational-force event rates for the 18-month period was correlated with CNC rates (Spearman's $r = 0.60$; $P < 0.0001$) (Table 1).

Association of at-fault CNCs with gravitational-force event rates

The final logistic regression/generalized estimating equations model includes the composite measure in the previous month and time since licensure (in half-years). Table 2 summarizes the numerical results of fitting this model. Accordingly, as the gravitational-force event rate increases so does CNC risk (odds ratio = 1.07, 95% confidence interval: 1.02, 1.12). Moreover, crash risk was greatest during the first 6-month period compared with the second and third 6-month periods. Figure 3 shows the estimated risk of having at least 1 CNC event in a month as a function of the gravitational-force event rate in the previous month, with a separate curve for each 6-month period.

Prediction of at-fault CNCs

Figure 4 shows a cross-validated ROC curve for the predictor based on the above logistic regression model. The AUC was estimated to be 0.76 (95% confidence interval: 0.71, 0.80).

Modeling considerations

The selection of the final model involved the following considerations.

Composite measure versus separate gravitational-force events. We considered a modification of the above model with the composite measure replaced by the 5 separate gravitational-force event rates. The regression coefficients that included the 5 terms separately were not significantly different ($P = 0.3415$), supporting our use of the composite measure.

How to define "prior." We considered different ways to define the period for calculating the composite measure of elevated gravitational-force events prior to the current month, including the previous month, 2 weeks, 1 week, 100 miles, and 1,000 miles driven. The previous month had the strongest association and best predictive performance.

Possible variability in the relation between CNCs and prior gravitational-force event rates over time. When added to the model, the indicator for CNC occurrence in the previous month was borderline significant ($P = 0.0167$), but the corresponding AUC for prediction was actually slightly smaller. For parsimony, the final model did not include the CNC occurrence in the previous month.

Possible interaction between the composite measure and time-since-licensure. When an interaction term was added to the regression model, it was not significant ($P = 0.2665$). Therefore, the odds ratio reported in Table 2 is applicable to each of the three 6-month periods. Figure 3 suggests that the

Table 2. Association (Odds Ratios) of Monthly Occurrence of At-Fault or Partial-Fault CNCs With the Composite Measure of Elevated Gravitational-Force Events/100 Miles[a] for the Previous Month and Time Since Licensure in Half-Years, Naturalistic Teenage Driving Study, 2006–2009

Covariate	OR	95% CI	P Value (2 sided)
Composite gravitational-force event rate	1.07	1.02, 1.12	0.0051
Half-year			
2 vs. 1	0.46	0.29, 0.72	0.0006
3 vs. 1	0.53	0.32, 0.89	0.0160

Abbreviations: CI, confidence interval; CNCs, crashes and near crashes; OR, odds ratio.

[a] One mile = 1.6 km.

risk difference was greatest in the first 6-month period. However, this may be an artifact due to the logistic model. A direct test of this hypothesis (i.e., a test for interactions under the identity link) was not clearly significant ($P = 0.07$), but this statistical trend suggests that the relation between the composite measure and CNC risk was somewhat stronger in the first time period.

DISCUSSION

It should surprise no one who drives a vehicle that high rates of elevated gravitational-force events increase the likelihood of a crash. Our findings provide an objective test of the Wahlberg (3) hypothesis that the likelihood of a crash is roughly equal to the recent rate of elevated gravitational-force events. We found a significant association between elevated gravitational-force event rates in the prior month and CNC rates in the following month, and the AUC indicated that elevated gravitational-force event rates were about 76% accurate in predicting the likelihood of a CNC within 1 month. Notably, the AUC of 0.76 is similar to the Framingham risk index for cardiovascular disease (13).

The study examined driving behavior over the first 18 months of licensure when crash risk is particularly high (5). Previous analyses of these data indicated that elevated gravitational-force rates of teenage drivers were 5 times higher than adult rates and did not decline significantly over the 18-month study period, although CNC rates did (8), suggesting that novice teenage drivers got better at driving in a risky manner without crashing. A nonsignificant trend suggested that the relation between risky driving behavior and the probability of a CNC was higher in the first 6-month period.

The study provides perhaps the best evidence so far reported that elevated gravitational-force event rates predict the likelihood of a CNC in the near future. The objective information on risk and CNCs from in-vehicle data recorders overcomes the limitations of previous research and enabled calculation of ROC/AUC. The analyses were adjusted for autocorrelation, and the final model was determined after substantial preliminary analyses that ruled out possible confounders and interaction effects.

Limitations of the study include a relatively small regional sample that may limit the generalization of the findings to the general population of teenage drivers. The data do not indicate that specific elevated gravitational-force events caused specific crashes, only that the pattern of gravitational-force event rates predicted the likelihood of a crash. The CNC experience during the first month of licensure, known to be extremely high (5–7), could not be included in the analyses given the absence of predictor data. To confirm the predictive
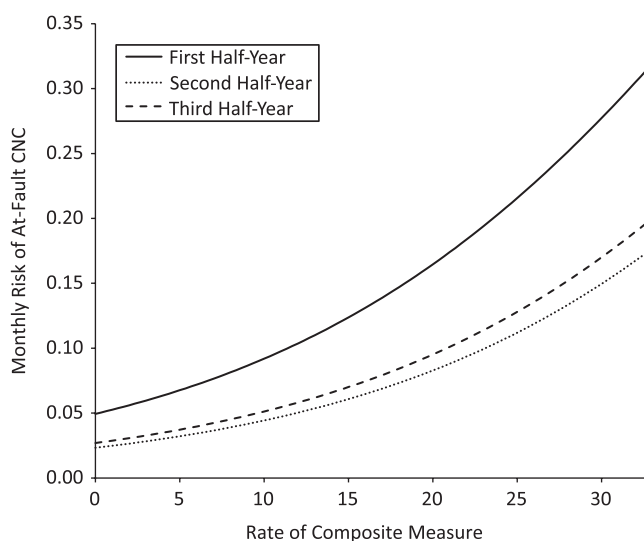

Figure 3. Estimated risk of having at least 1 at-fault crash and near-crash (CNC) event in a month as a function of the composite measure of elevated gravitational-force events (per 100 miles) in the previous month and time since licensure (in half-years), Naturalistic Teenage Driving Study, Blacksburg, Virginia, 2006–2009. One mile = 1.6 km.
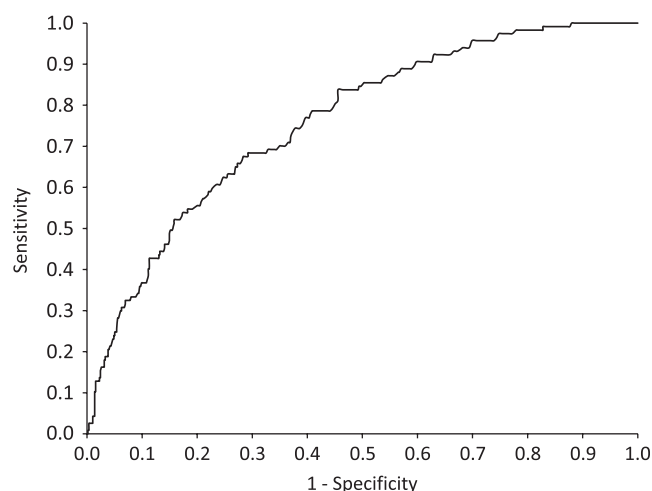

Figure 4. Leave-one-out cross-validation receiver operating characteristic curve and area under the curve (AUC = 0.7601) for predicting at least 1 at-fault crash and near-crash event in a month as a function of the elevated gravitational-force event rates (per 100 miles) in the previous month, Naturalistic Teenage Driving Study, Blacksburg, Virginia, 2006–2009. One mile = 1.6 km.

validity of the relation of prior elevated gravitational-force event rates and CNCs, investigators must replicate the method with similar data collected from other populations.

REFERENCES

1. Bagdadi O, Várhelyi A. Jerky driving—an indicator of accident proneness? *Accid Anal Prev.* 2011;43(4):1359–1363.
2. Elvik R. Laws of accident causation. *Accid Anal Prev.* 2006; 38(4):742–747.
3. Wahlberg AE. Aggregation of driver acceleration behavior data: effects on stability and accident prediction. *Safety Sci.* 2007;45(4):487–500.
4. Klauer SG, Dingus TA, Neale VL, et al. *Comparing Real-World Behaviors of Drivers With High Versus Low Rates of Crashes and Near-Crashes.* Washington, DC: National Highway Traffic Safety Administration; 2009. (DOT publication no. HS 811 091).
5. Williams AF. Teenage drivers: patterns of risk. *J Safety Res.* 2003;34(1):5–15.
6. National Highway Traffic Safety Administration. *Traffic Safety Facts 2008 Data Overview.* Washington, DC: National Highway Traffic Safety Administration; 2009. (DOT publication no. HS 811 162).
7. Twisk DA, Stacey C. Trends in young driver risk and countermeasures in European countries. *J Safety Res.* 2007; 38(2):245–257.
8. Simons-Morton BG, Ouimet MC, Zhang Z, et al. Crash and risky driving involvement among novice adolescent drivers and their parents. *Am J Public Health.* 2011;101(12): 2362–2367.
9. Guo F, Klauer SG, Hankey JM, et al. Near crashes as crash surrogate for naturalistic driving studies. In: *Highway Safety Data, Analysis, and Evaluation 2010.* Vol 1. Washington, DC: Transportation Research Board; 2010:66–74.
10. Liang KY, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986;73(1):13–22.
11. Zhou XH, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine.* New York, NY: John Wiley & Sons; 2002.
12. Picard R, Cook D. Cross-validation of regression models. *J Am Stat Assoc.* 1984;79(387):575–583.
13. Zomer E, Owen A, Magliano DJ, et al. Validation of two Framingham cardiovascular risk prediction algorithms in an Australian population: the 'old' versus the 'new' Framingham equation. *Eur J Cardiovasc Prev Rehabil.* 2011;18(1): 115–120.