The Interface Between Statistical Research and Teenage Driving: What Statistics Can Teach Us About How Our Kids Drive and Visa-Versa

Paul S. Albert

Biostatistics and Bioinformatics Branch
Division of Intramural Population Health Research
Eunice Kennedy Shriver National Institute of Child Health and Human Development
National Institutes of Health, DHHS

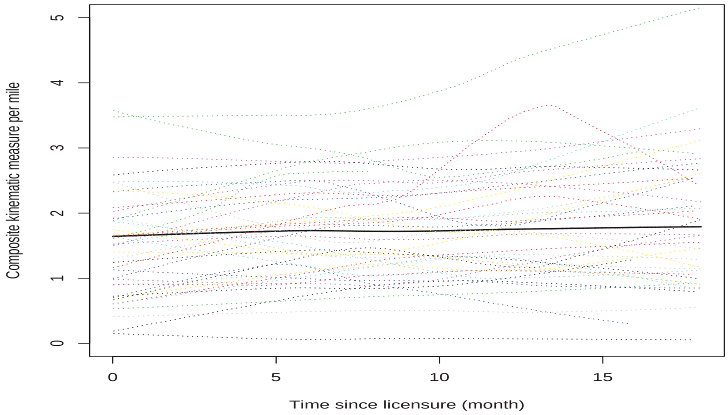
Fourth International Naturalistic Driving Research Symposium

Scientific Questions

- How does risky driving behavior measured by g-force events vary by condition and over time?
  - Do composite g-force events change over time?
  - Do trip-specific covariates (e.g. adult passengers, night driving, etc.) effect g-force events?
  - What are the sources of variation in g-force events?
  - What is the serial dependence in g-force events?
- How do g-force events relate to teenage accidents?
  - Can we predict actual or near crashes from g-force events?

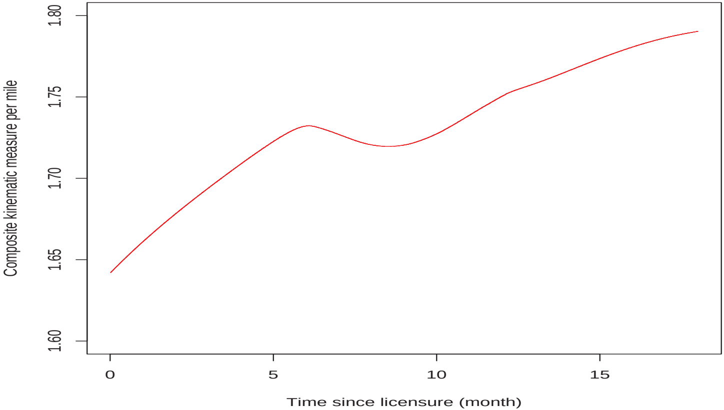
Outline

- Exploring features of the data
- Random process and marginal modeling of LONGitudinal counts:
  - A hierarchical Poisson regression modeling approach (Kim, Chen, Zhang, Simons-Morton, Albert, 2013 JASA)
  - Marginal analysis of longitudinal counts data in long sequences (Zhang, Albert, Simons-Morton, 2012 AOAS)
- Joint models of kinematic measurements and crashes for prediction
  - Ordinal latent variable models and their application in the study of newly licensed teenage drivers (Jackson, Albert, Zhang, Simons-Morton, 2013 JRSS-C)
  - A two-state mixed hidden Markov model for risky teenage driving behavior (Jackson, Albert, Zhang, In press at AOAS).
- Discussion
  - interesting problems?



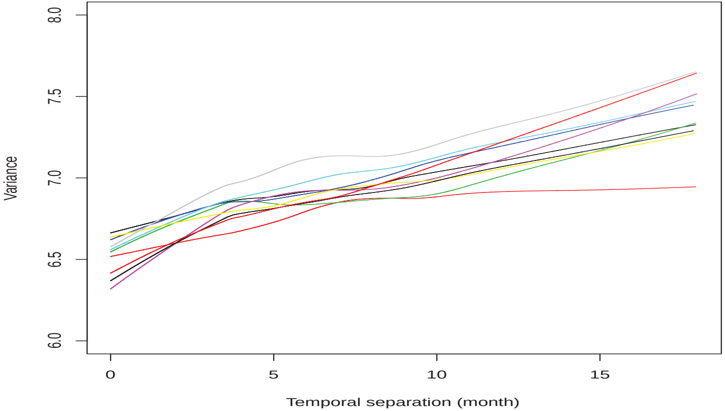
Individually Smoothed Curves

Exploring the Data (Continued)



Mean Trajectory

Exploring the Data (Continued)



Lowess smoothed empirical variograms for the composite kinematic events based on 10 random pairings with each observation in the dataset randomly paired with another on the same individual.

A Hierarchical Model

We assume the hierarchical Poisson regression models as follow:

$$y_{ij} \sim \text{Poisson}\left\{ m_{ij} \exp\left(g(t_{ij}) + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \tau_i + \gamma_{ij} + \epsilon_{ij} \right) \right\}$$

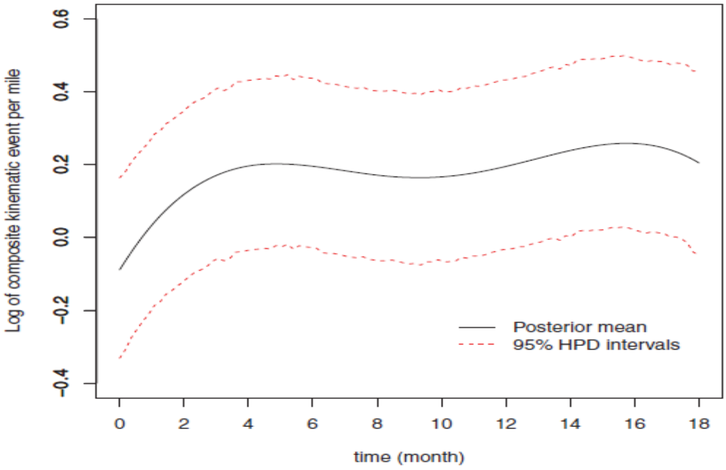
where

- $g(t_{ij})$ is a polynomial regression spline of order p with k knots.
- $\tau_i \sim N(0, \sigma_\tau^{*2})$: a random effect for subject
- $\gamma_{ij} \sim N(0, \sigma_\gamma^{*2})$: a random effect for overdispersion
- $\epsilon_{ij} \sim N\left(0, \sigma_\eta^{*2}\left(1 - \rho^{2d_{ij}}\right)\right)$ with $\rho = \exp(-\theta)$ and $d_{ij} = |t_{ij} - t_{i,j-1}|$: a random effect with serial correlation among trips

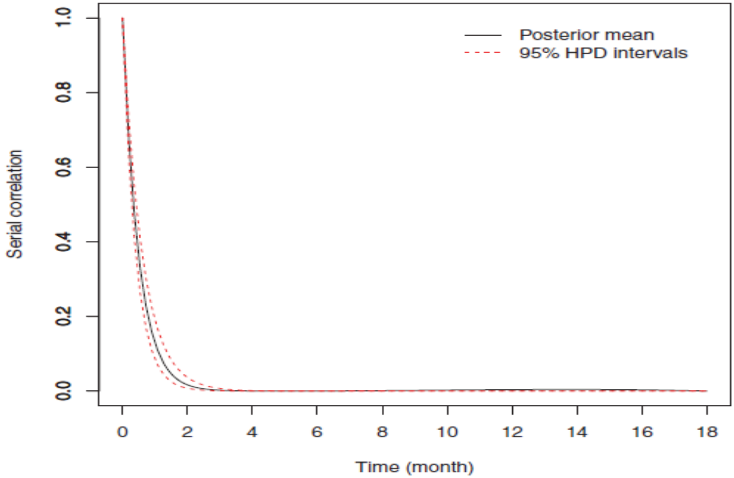
Posterior Estimates Under the Full Model

Variable	Posterior Mean	Posterior SD	95% HPD Interval
passenger	-0.181	0.006	(-0.194, -0.168)
day/night	-0.193	0.006	(-0.204, -0.182)
risky friend	0.406	0.168	(0.072, 0.729)
σ_{τ}^{*2}	0.287	0.070	(0.165, 0.423)
σ_{γ}^{*2}	0.269	0.003	(0.263, 0.275)
σ_{η}^{*2}	0.125	0.006	(0.113, 0.137)
θ	36.824	3.709	(29.834, 44.260)

Estimated Log-longitudinal Trajectory



Serial Correlation (Model Based)

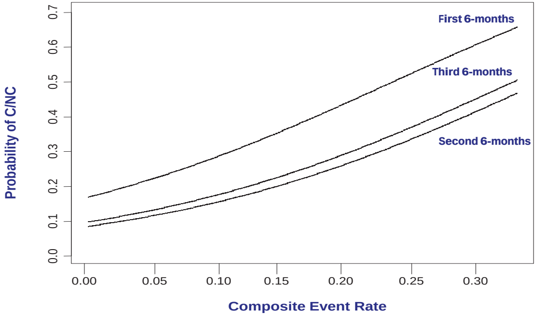


Kinematic measures and their correlation with C/NCs

Category	*g-force*	*Frequency*	*% total events*	*Correlation with CNCOs†*
Rapid starts	> 0.35	8747	39.6	0.28
Hard stops	$\leqslant -0.45$	4228	19.1	0.76
Hard left turns	$\leqslant -0.05$	4563	20.6	0.53
Hard right turns	$\geqslant 0.05$	3185	14.4	0.62
Yaw	6^0 in 3s	1367	6.2	0.46
Total		22090	100	0.60

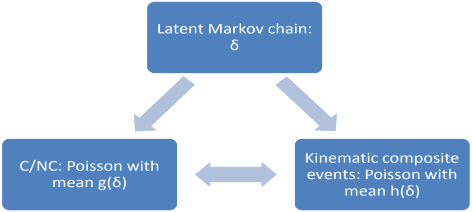
†Correlation computed between the CNCO and elevated *g*-force events based on monthly rates.

GEE With Logistic Regression
Prediction of C/NC by Period



Simons-Morton et al., American Journal of Epidemiology, 2012

Joint Model for C/NC and Kinematics: A Hidden Markov Modeling Approach



Hidden Markov Model:Prediction

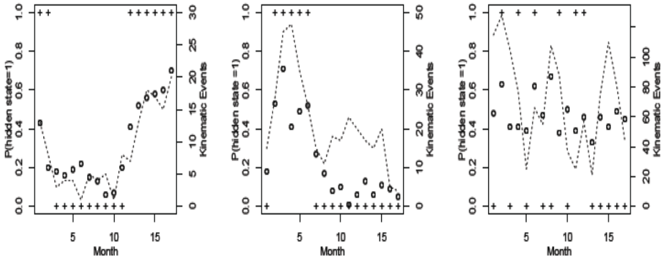


Figure 2: Predicted value of the hidden state given the observed data for three drivers. The (o) indicates the probability of being in state 1 (poor driving),(+) indicates a crash/near crash event and the dotted line indicates the composite kinematic measure.

Hidden Markov Model:Prediction

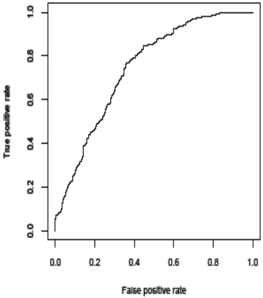


Figure 1: ROC curve for the mixed hidden Markov model based one 'one-step ahead' predictions (area under the curve = 0.74).

Summary

- Exciting opportunities for collaborative work with research statistical scientists
  - Understanding variation in kinematic measurements
  - Developing dynamic predictors of crashes
- Future research
  - Identify subgroups of teenagers that are at extreme risk: Tree-based approaches
  - Understanding effect of a C/NC on subsequent kinematic dynamics: Recurrent events
  - Cost-effective and efficient designs for large scale studies

References

1. Jackson, J.C., Albert, P.S., Zhang Z., and Simons-Morton, B. Ordinal latent variable models and their application in the study of newly licensed teenage drivers. Journal of the Royal Statistical Socieity- Series C 62, 435-450, 2013.
2. Jackson, J.C., Albert, P.S., and Zhang Z. A two-state mixed hidden Markov model for risky teenage driving behavior. In Press at Annals of Applied Statistics.
3. Kim, S., Chen, Z., Zhang, Z., Simons-Morton, B., and Albert, P.S. Bayesian hierarchical Poisson regression models: an application to a driving study with kinematics events. The Journal of the American Statistical Association 108, 494-503,2013.
4. Zhang, Z., Albert, P.S., and Simons-Morton, B. Marginal analysis of longitudinal count data in long sequences: methods and applications to a driving study. Annals of Applied Statistics 6, 27-54, 2012