VARIABLE INTERPRETATIONS

- Most traffic crashes in urban areas of Chinese cities occur at signalized intersections.
- Due to the small distance between signalized intersections (averaging 199meters) in urban areas of Shanghai, it is essential to account for the within-corridor correlations among intersections.
- Speeding along corridors were anticipated to be important predictors of crashes.
- The purpose of this study is to develop a Bayesian hierarchical crash prediction model for signalized intersections in Shanghai that can deal with within-corridor correlations among intersections, and can properly evaluate the risk factors, especially the speed features for intersection crashes.

DATA PREPARATION

- Sampling: 195 signalized intersections along 22 different corridors in urban areas were sampled.

Data Structure: The data structure used in this study can be viewed as a two-level hierarchy with level 1 being the intersection level, and level 2 being the corridor level.

Intersection and Corridor Related Variables: A total of 17 intersection related variables and 6 corridor related variables were collected for model development.

Speed Features: The FCD of one week between the hours of 13:00-14:00 was used to determine average speed and speed variance for each corridor. The positions of taxi sample equipped with GPS devices were recorded every ten seconds. Taxi samples which passed through the study corridor were chosen.

- The speed $V$ for taxi sample $i$ can be expressed as $V = \frac{\sum_{j=1}^{m} d_j}{\sum_{j=1}^{m} t_j}$.
- The average speed $V$ and speed variance $D(V)$ for a corridor:
  
  $\sum_{j=1}^{m} \frac{d_j}{t_j} = D(V) = \frac{1}{n-1} \sum_{j=1}^{m} \left( V_j - \bar{V} \right)^2$

METHODOLOGY

- Poisson Models: can accommodate the nonnegative, random and discrete features of crash occurrence, but the over-dispersion in crash data violates the assumption of the Poisson distribution that the variance of the crash data is constrained to be equal to the mean.
- Negative Binominal Models: can deal with the over-dispersion in crash data by introducing an error term, but are based on the plausible assumption that observations are independent from each other.
- Generalized Estimating Equations: account for the dependency among observations, but cannot reflect the discrepancy of different intersection groups.
- Hierarchical Models: can offer an improved method to address multi-level data structures. The merits of hierarchical models are the ability to handle correlated data, and consequently, the ability to include covariates at the intersection and corridor levels, thus allowing the effects of intersection and corridor variables to be independently evaluated.

MODELING RESULTS

- FIXED EFFECTS: For each corridor. The positions of taxi sample equipped with GPS devices were recorded every ten seconds. Taxi samples which passed through the study corridor were chosen.
- Deviance Information Criterion (DIC): can be used to compare complex models, because it offers a Bayesian measure of model fitting and complexity.
- Deviance Information Criterion (DIC) = $D(D(\theta) + p_D)$

CONCLUSIONS

- The HNB model which accommodates both individual and corridor level dispersion performed better than other models considered in this study. However, the corridor correlation between signalized intersections.
- The NB model which accommodates both individual and corridor level dispersion performed better than other models considered in this study.
- Deviance Information Criterion (DIC) = $D(D(\theta) + p_D)$