NHTSA Mass-Size-Safety Symposium

Paul E. Green

Washington DC
February 25, 2011

UNIVERSITY OF MICHIGAN
TRANSPORTATION RESEARCH INSTITUTE

Overview

- Background on Mass-Size-Safety
- Data Sources
- Some Current Approaches Using Statistical Models
- Multicollinearity
- Some Suggestions
- Induced Exposure
- The Future

Background

- NHTSA selected footprint attribute upon which to base CAFE standards for MYs 2012-2016.
- These standards are likely to result in weight reductions in new cars and light trucks.
- Government would like to estimate effect of new CAFE standards on safety (injuries, fatalities).
- A number of studies have been conducted demonstrating an association between fatality rates and curb weight, track width, and wheelbase.
- Many of the studies are not consistent.

Background

- Some studies report a decrease in fatalities with vehicle weight reduction, others report an increase.
- Other studies suggest stiffness, frontal height, and vehicle design are better related to fatality rates than weight.
- The various studies are generally based on different underlying assumptions.
- The assumptions include different choices about variables, databases, and statistical models.
- Investigators tend to have different backgrounds and philosophies.

Notes for Consideration

- Analyses have been based on historical data.
- Innovations in materials that provide strength at lighter weights and advances in occupant protection systems may change relationships.
- There have been recent advances in active safety technologies (ESC, ACC, LDW, etc.)
- Important that methods for estimating future vehicle safety take into account advances in technology.

Data Sources

- **Crash Data**
  - FARS – Census file of fatal involvements
  - CDS – Survey , severity threshold, crash investigations
  - State Data – Induced exposure involvements, non-culpable vehicles in two-vehicle crashes
  - Other Sources – curb weight, track width, wheelbase
- **Exposure Data**
  - VMT – Vehicle miles travelled (not recorded at level required)
  - Vehicle registrations
- **Some Databases Created Appear to be Impressive**

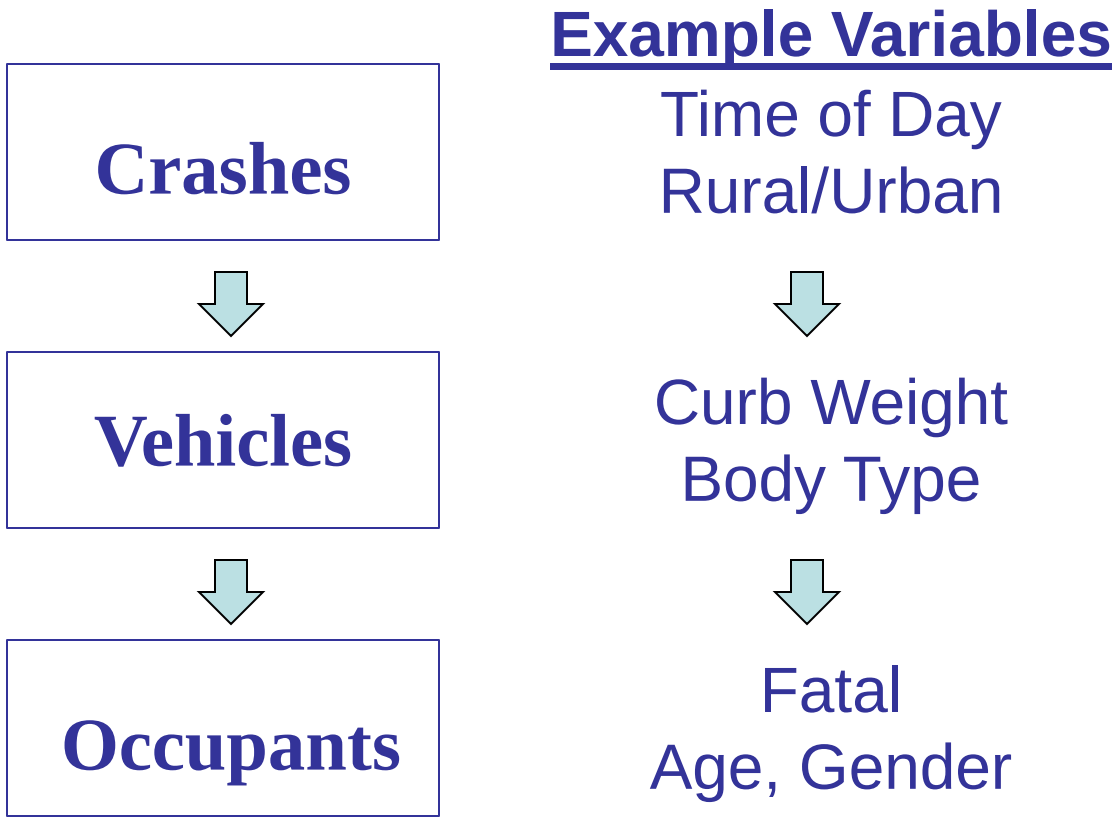
Examples of Variables Under Investigation

Some variables are continuous, others categorical

- **Vehicle** – curb weight, track width, wheelbase
- **Vehicle type** – pass cars, minivans, SUVs, light trucks
- **Driver** – age, gender, belt use, alcohol use
- **Roadway** – straight/curve, limited access/major artery/other, intersection, speed limit
- **Environment** – rural/urban, dry/wet, day/night
- **Crash type** – single-vehicle, head-on, rear-end, sideswipe, crossing paths
- **Crash severity** – fatal, injury, property damage

Crash Data are Hierarchical

Crash data are generally arranged into Accident, Vehicle, Person, etc files



Can Regression Models be Used to Relate Vehicle Mass and Size to Fatality Risk?

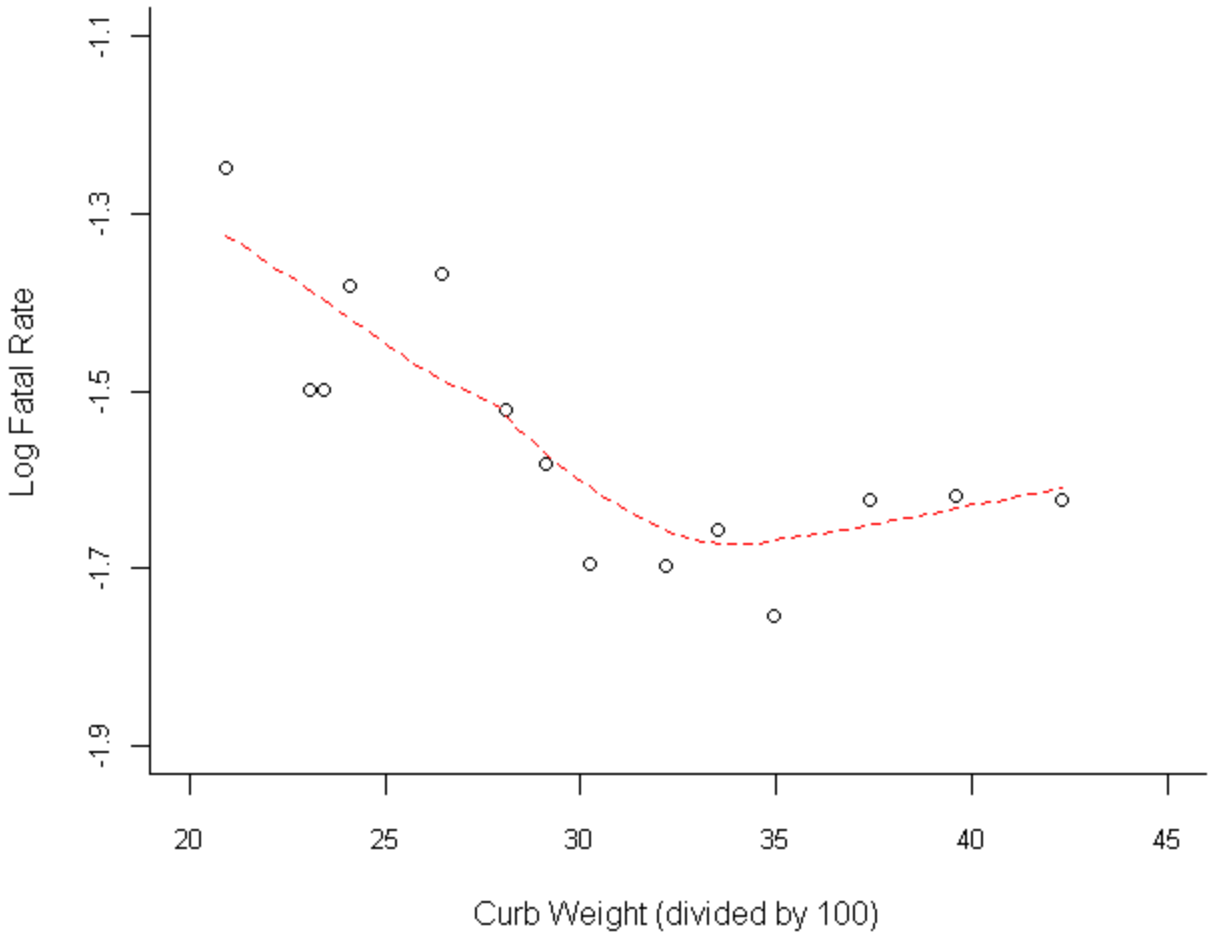
All Models are Wrong, Some are Better than Others, Some are Useful

We seek models that describe the truth, even though we know we cannot describe it exactly.

Claim: Most of us would likely say that we know a good model when we find one.

Applied statistics is an art form.

Scatter Plot of Log Fatal Rate by Curb Weight



Kahane 2003

Traditional Exposure-Based Risk Models Used for Analyzing Fatality Rates

- **Poisson log-linear models (generally too simple)**
- **Negative binomial log-linear models**
- **Weighted least squares log-linear models**
- **Random effects models**

These models generally require aggregated data

Disaggregate Logistic Regression

- Some reports used disaggregate logistic regression to model fatality risk as a function of vehicle mass and size.
- Appears data were collected at the vehicle level.
- Assumes observations are independent.
- Can be used as an alternative to one of the more traditional exposure-based risk models.
- Likelihood-based tests can overstate significance.

Multicollinearity

- Historical data – curb weight, track width, wheelbase highly correlated.
- Can lead to unstable estimation.
- Parameter estimates can change sign.
- Parameter estimates can change magnitude.
- Centering variables can help.
- Our recommendation is to not include highly correlated variables in the same model.

Suggestion

- Perform a matched analysis.
- Match on footprint and possibly age and gender.
- Matched variables are controlled and not fit.

				Matched Variables			Other Variables	
Stratum	Fatal	Vehicle Registration Years	Curb Weight	Footprint	Driver Age	Driver Gender	Nite	Rural
1	1	1	3000	40	25	M	Day	Rural
1	0	295	2500	41	24	M	Nite	Rural
2	1	1	2850	45	37	F	Nite	Urban
2	0	300	3500	44	38	F	Nite	Urban
3	1	1	2100	39	46	M	Day	Urban
3	0	270	2800	39	45	M	Day	Urban

Why Match?

- **Matching is a tool specifically designed to control for confounders.**
  - Footprint
  - Age, gender
- **Results in more efficient estimation.**
  - More efficient when confounder is associated with both the response variable and the predictor of interest.
  - Footprint is associated with fatality risk and curb weight.
- **Can focus on the effects of curb weight while holding footprint constant.**

Examination of Residuals

- Can be used to detect outlying observations.
- Some issues related to 2-door versus 4-door cars.
  - "Sporty" or "muscle cars" (trackwidth, wheelbase)
- Large residuals could alert the analyst to poorly fitting observations.
- Detection of outliers could lead to new research questions.

Induced Exposure

- It is recognized that there are no good sources of exposure (VMT) recorded at the level needed for studies such as these.
- Some studies use induced exposure.
- Induced exposure vehicles are non-culpable vehicles involved in multiple-vehicle crashes.
- Some concern about the effects of induced exposure on results (bias, sensitivity).

The Future

- Predicting the future effects of size/weight on fatality risk using historical data is very difficult.
- Some trends have already been discovered with respect to active safety technologies (ESC).
- Similarly, effects of size/weight should become evident as newer data becomes available.
- Simulation could be a valuable tool in certain controlled settings.
- Simulation techniques and data analytic methods are both useful.

Thank You !