WHEN MORE IS TOO MUCH: AGGREGATION BIASES IN SAFETY RESEARCH

Gary A. Davis
University of Minnesota

88th Annual TRB Meeting
Session 350
January 12, 2009

WHAT ARE ‘AGGREGATION BIASES’?

Many transportation-related phenomena result from actions of individuals

- Mode/Route Choice
- Road Crashes

Data are often available (or conveniently processed) as aggregates of individuals

- Census Tracts
- Highway Sections

Aggregation biases occur when analyses using aggregate data produce misleading conclusions about individual behavior or actions

ECOLOGICAL FALLACY

Term introduced by Robinson (1950)

"The ecological fallacy consists in thinking that relationships observed for groups necessarily hold for individuals…" -David Freeman

In safety research, ecological fallacies can arise, e.g., when associations between crash rate and average speed are used to draw conclusions concerning speed and individual crash risk

SIMPSON'S PARADOX

Statistical associations observed at one level of aggregation can disappear (or be reversed) at a different level

Judea Pearl explains this as the result of neglecting underlying causal mechanisms

SIMPSON'S PARADOX
Classic Example

Hypothetical Results from an Observational Study of Drug Effectiveness

Males

	Cured	~Cured	Recovery Rate
• Drug	18	12	60%
• No Drug	7	3	70%

Females

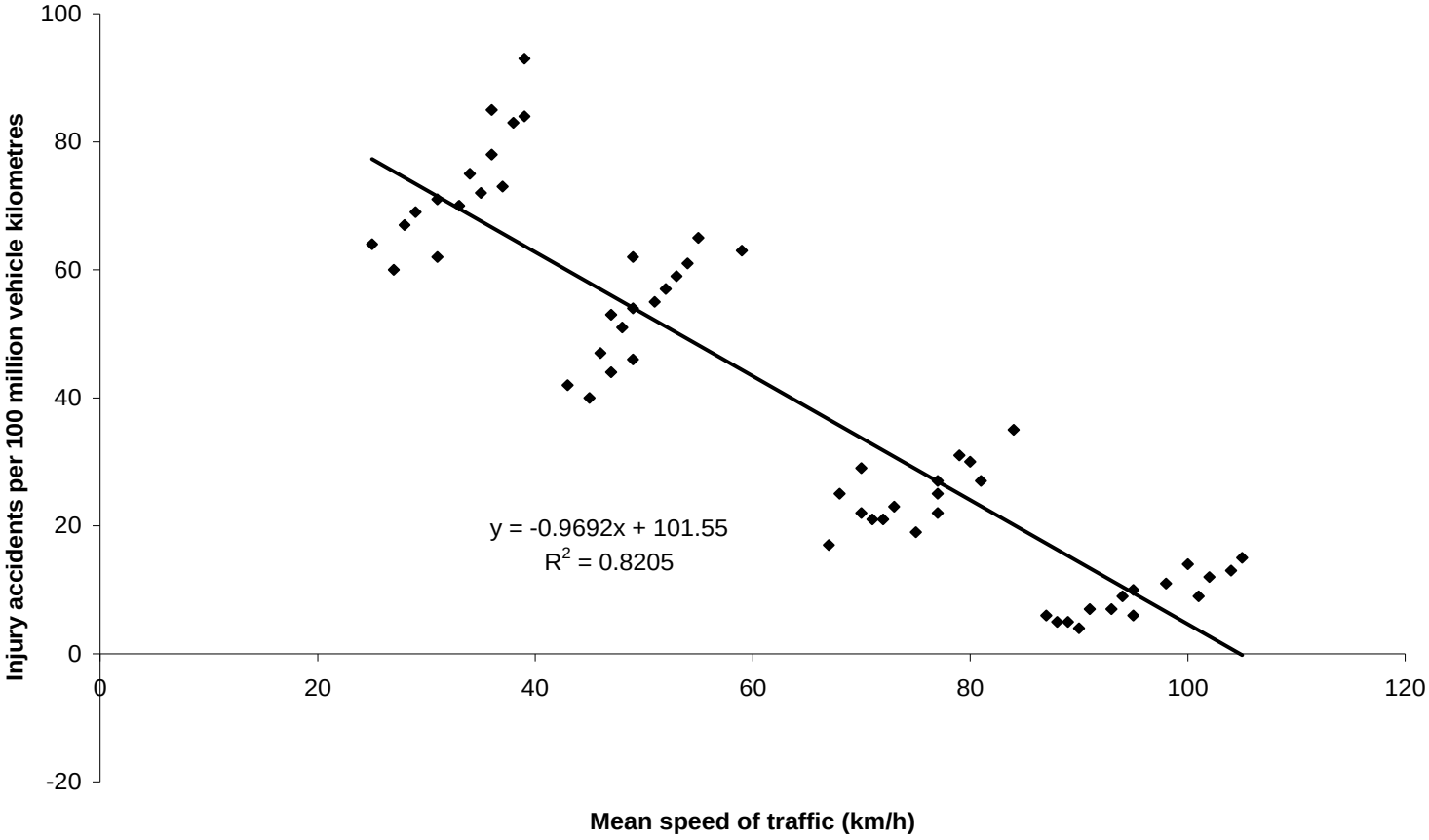
	Cured	~Cured	Recovery Rate
• Drug	2	8	20%
• No Drug	9	21	30%

Now suppose the data are aggregated

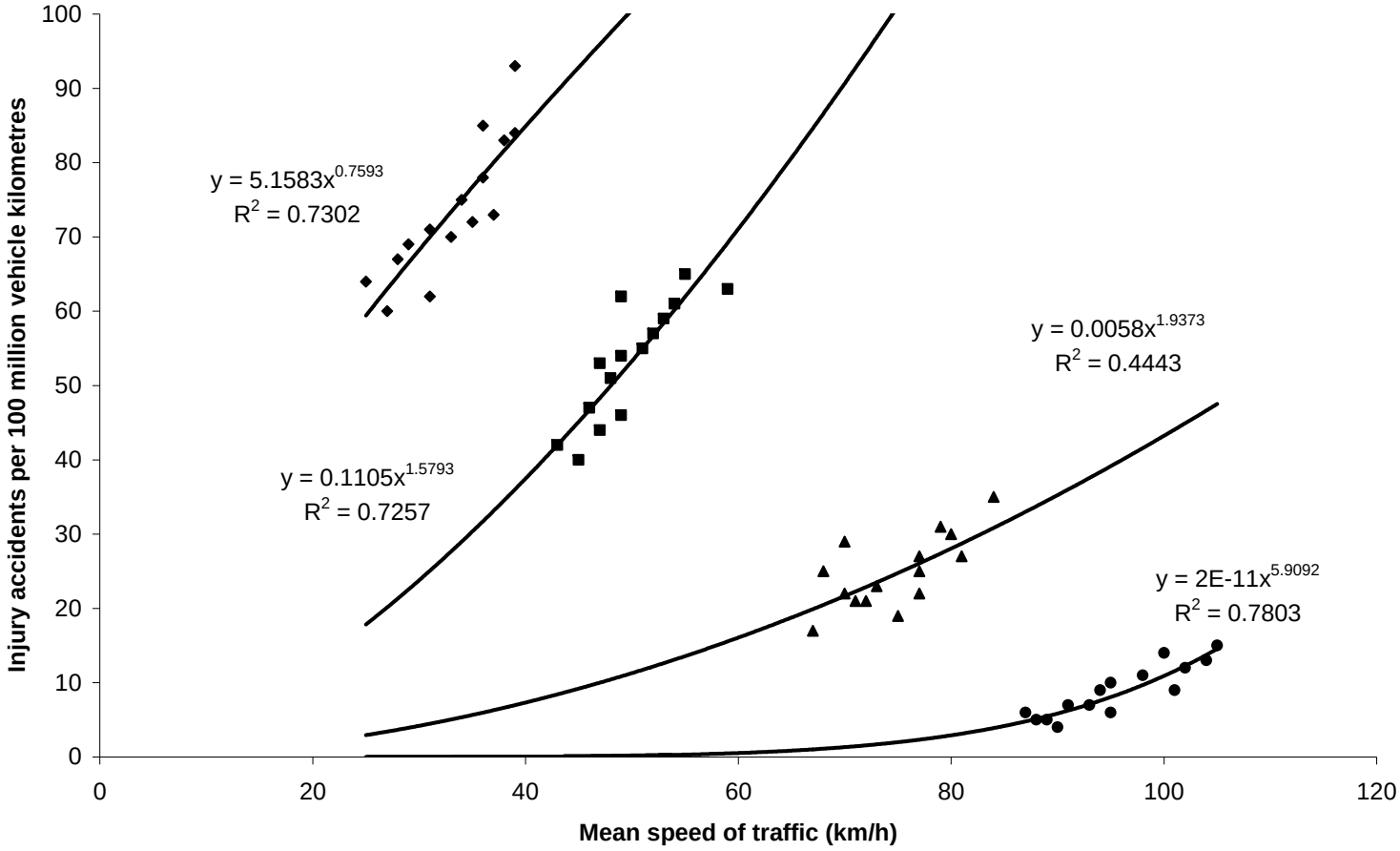
Combined Data

	Cured	~ Cured	Recovery Rate
• Drug	20	20	50%
• No Drug	16	24	40%

NON-HYPOTHETICAL EXAMPLE (from R. Elvik)

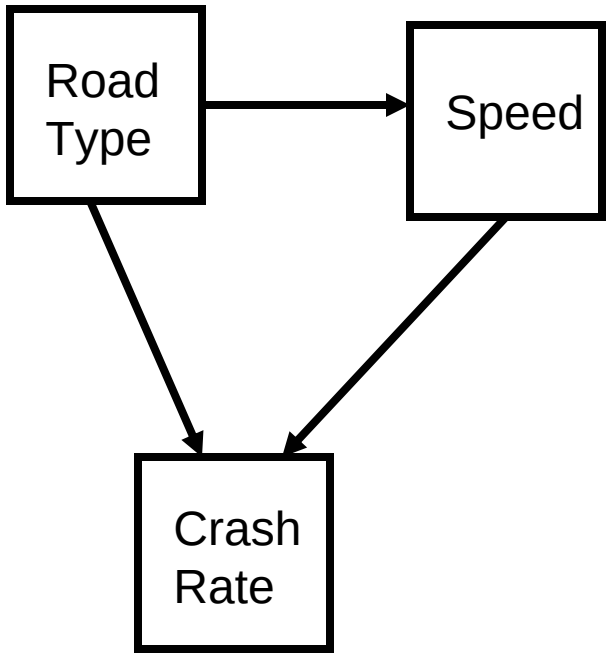


ELVIK'S DISAGGREGATED ANALYSIS

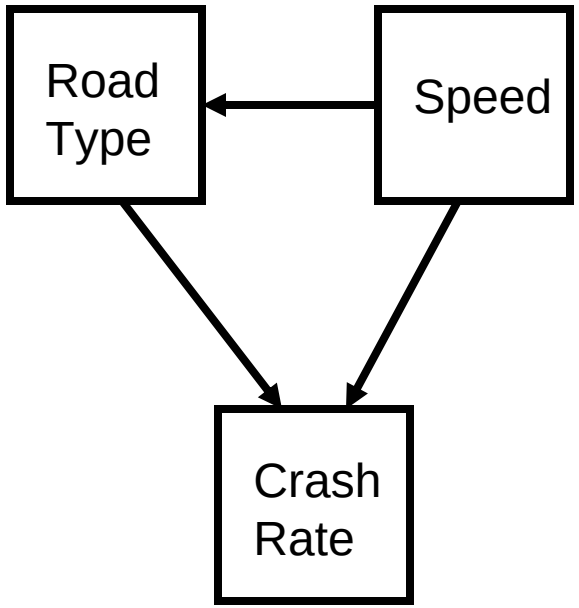


JUDEA PEARL:
Correct Interpretation Depends on Underlying Causal Process

Disaggregate!



Aggregate!



SOLOMON'S CURVE

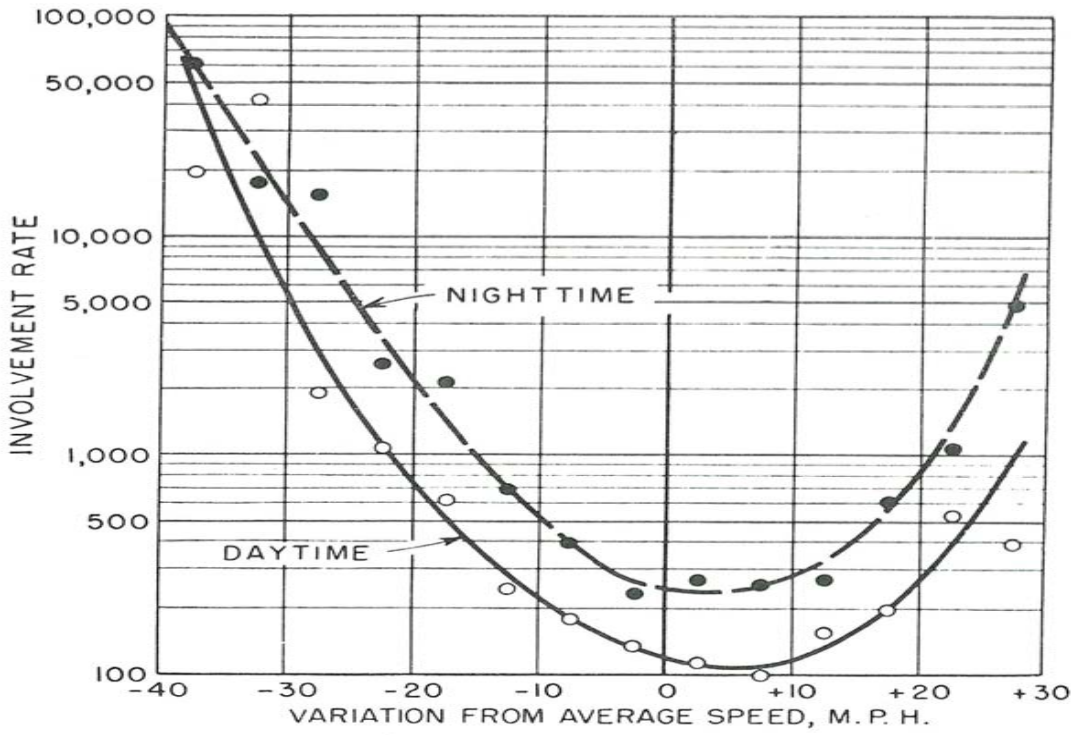


Figure 7.—Involvement rate by variation from average speed on study section, day and night.

HOW WAS CURVE GENERATED?

Involvement Rate = P[v|crash]/P[v]

P[v|crash] estimated from crash reports
(including vehicles at intersections)

P[v] estimated from spot-speeds
(did not include vehicles at intersection)

SOLOMON'S CURVE A SIMPSON'S PARADOX?

Hypothetical 'Solomon' Data

Speed(v)	P[crash|v]	P[v]	P[v|crash]	IR
Intersection-Related				
Low	.01	.8	.47	.59
Medium	.01	.1	.059	.59
High	.08	.1	.47	4.7
Road Departure				
Low	.01	.1	.059	.59
Medium	.01	.8	.47	.59
High	.08	.1	.47	4.7

COMBINING CRASH TYPES

P[intersection|crash]=P[road-depart|crash]=0.5

Speed(v)	P[v]	P[v|crash]	IR
Low	.1	.265	2.6
Medium	.8	.265	.33
High	.1	.47	4.7

COMMENTS FROM ELSEWHERE

"Aggregate data are often easier to obtain than data on individuals, and may offer valuable clues about individual behavior. Ecological inferences will therefore continue to be made. The problems of confounding and aggregation bias, however, are unlikely to be resolved in the proximate future." (Freedman, 1999, after reviewing examples of ecological bias in social science)

"As with other observational studies, ecologic studies can give useful results if biases such as those discussed here can be ruled out. Nevertheless, bias evaluation can be especially difficult in ecologic studies of geographic regions because of the many potentially interacting covariates that may differ across regions. When biases cannot be ruled out with available data, ***further exploration will require individual-level studies."*** (Greenland and Robins, 1994, after reviewing ecologic biases in epidemiology)

CONCLUSION

2 Views on the Nature of Road Accidents:

1. Accidents result from the workings of '**chance-setups':** physical systems that produce stable relative frequencies but unpredictable individual outcomes. Uncertainty results from inherent 'chanciness.'

2. Accidents result from the workings of (approximately) **deterministic mechanisms**. Uncertainty arises from ignorance concerning the operation of or inputs to these mechanisms.

My Opinion:

The **statistical regularities** seen in observational safety studies **have no independent existence**. Rather, they are they outcome of aggregating particular types and frequencies of individual accident mechanisms.

Progress in understanding the causes of road accidents **requires** investigation at **low levels of aggregation.**