Approaches to Assessing + Communicating About AV Safety

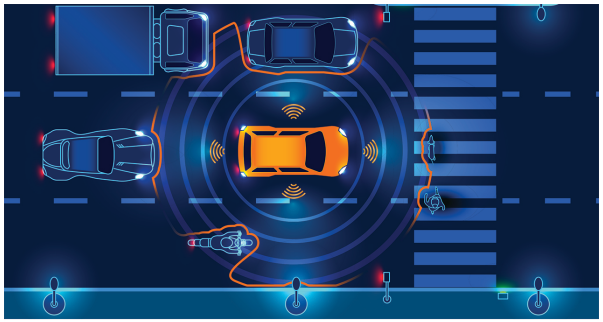
Marjory S. Blumenthal

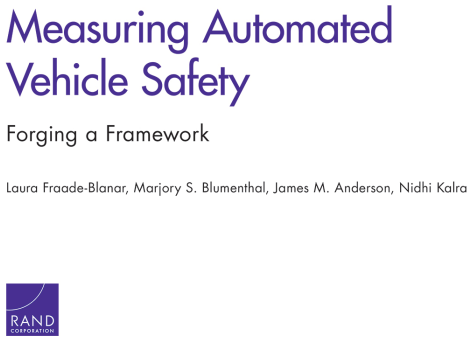
Senior Policy Researcher

RAND Corporation

Builds on *Measuring Automated Vehicle Safety: Forging a Framework*

- Combine measurements into a framework
  - In *settings* (simulation, closed courses, public roads)
  - At *stages* (development, demonstration, deployment)
- Leading (pre-crash) measures are key
- *Roadmanship* concept: Drive safely without creating—and respond well to—hazards





RAND

Issues for Assessing + Communicating

- Three principal approaches:
  - Measurement
  - Processes
  - Thresholds
- Asymmetric information context—developers know the most
- Diverse stakeholders and audiences

RAND

Measurement

- Gold standard but elusive
  - Lack of lagging measures
- Reliance on immature leading measures
  - *Roadmanship* concept is implicit in ongoing efforts
  - Candidate-measure scorecard—consider ability to validate, applicability to different crash types, incentive effects, …
- Nonuniformity frustrates stakeholders

RAND

Leading Measure Challenge Example

Braking Event	Danger Present	Danger Absent
Hard braking occurred	A: appropriate reaction	B: false positive
No hard braking	C: false negative (crash)	D: appropriate avoidance

Processes

- Compensate for measurement weaknesses
- Indicators of developer attention to safety
  - AV response to unanticipated circumstances?
- Different forms:
  - Compliance with regulation (*limited*)
  - Implementation of technical standards (*growing*)
- Cross-cutting: safety cases and safety culture
- May not be transparent—internal, proprietary

RAND

Thresholds—Quantitative *or* Qualitative

1) Human driving comparison
   - Intuitive, sought-after
   - Average v. better or “safe” human driver
   - ODD-specific (but data dearth)
2) Automated driving performance
   - Driving test +/-
   - ALARA/ALARP
     - “Positive trust balance”
3) Absolute goal
   - Vision Zero +/-
   - GAMAB/MEM
     - Life is full of risk

RAND

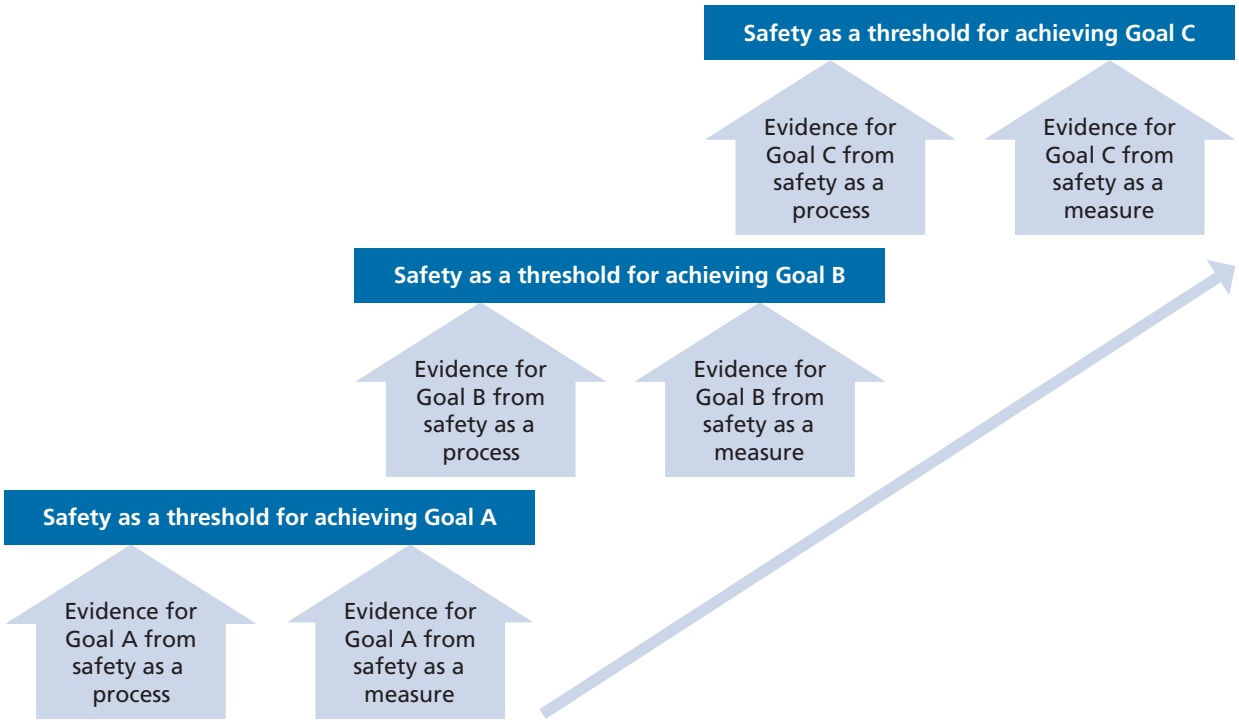
Threshold Comparison

Threshold	Conceptually	Functionally
Human drivers	Strong	Weak
ADS technology performance	In development	In development
Absolute goals	Weak	Strong

RAND

No One Assessment Approach Suffices

- Approaches complement, support, interact
- Continuing progress—staircase-like



Communicating About Safety

- Broad agreement on challenge
  - Public trust
- Risk perception
  - Heuristics, biases
  - Perception of control
  - Personal experience
  - Quantitative savvy
- Affect heuristic
  - Perceptions ⇄ Emotions
    - Exposure and habituation (e.g., ADAS)
- Experience elsewhere
  - Human error > machine error
  - Assume safety thresholds met

RAND

Communicating Approaches

Process *or* Measure		Threshold		Statement
AV hard-braking rate	is	below 1 per million VMT	as	communicated by AV developer
AV fatal crash rate per 100 million VMT	is	lower than that of the average human driver	as	communicated by government statistics
AV safety case demonstrates meeting safety standards	showing that	the technology is as safe as possible	as	communicated by a safety advocacy group

RAND

American Life Panel Survey

- Standing consumer panel
- Responses to patterns of info from 8 different sources
  - Relative influence

Source of evidence	Evidence shows that AVs are safe
Average AV crash rate	
Average near-miss crash rate	
Federal vehicle requirements	
Federal government official position	
State or local government official position	
AV company's official position	No information
Safety advocacy group's official position	
Friends or family members	

Explicit Rankings v. Implicit Influence

	Safety Message Source	
	Regression Coefficient (implicit)	Rank Order (explicit)
1	AV crash rate	AV crash rate
2	State or local government position	Safety advocacy group position
3	Federal government position	Federal vehicle requirements
4	AV near-miss rate	AV near-miss rate
5	Federal vehicle requirements	Federal government position
6	Safety advocacy group	State or local government position
7	Friends and family members	Friends and family members
8	AV company position	AV company position

NOTE: Order of sources measured implicitly determined by standardized regression coefficients (see Table A.1) from the social judgment analysis. Order of sources measured explicitly determined by mean ranking from the rank-order task.

Special Survey Showed Influence of, Preference for Info Sources

- Most compelling
  - AV crash rates (but elusive as a measure)
  - Info from state and local government
  - Information from the federal government
  - AV near-miss rates
- Most effective
  - Data-driven, immediately understandable and relevant (AV crash rates)
  - State, local, and federal government preferred to companies, friends + family

RAND

No Single Message Suffices

- Populations have different needs
  - Different perceptions of risks, benefits
  - Different views of source credibility
- Promote AV benefits, don't talk only about risks and costs
- Use simple, data-driven statements from trusted sources

RAND

Safe Enough Recommendations

Developers

- Use a mix of approaches for safety assessment
- Continue to advance leading measures, incl. roadmanship
- Collaborate on templates for publicly assessible versions of safety cases
- Bring AVs into communities

Government

- Support research into (and data about) human drivers to enable good ODD-specific comparisons
- Support research into safety assessment options, especially measurements

Questions?

Marjory@RAND.org