

# Update

---

## SHRP2 Naturalistic Driving Study

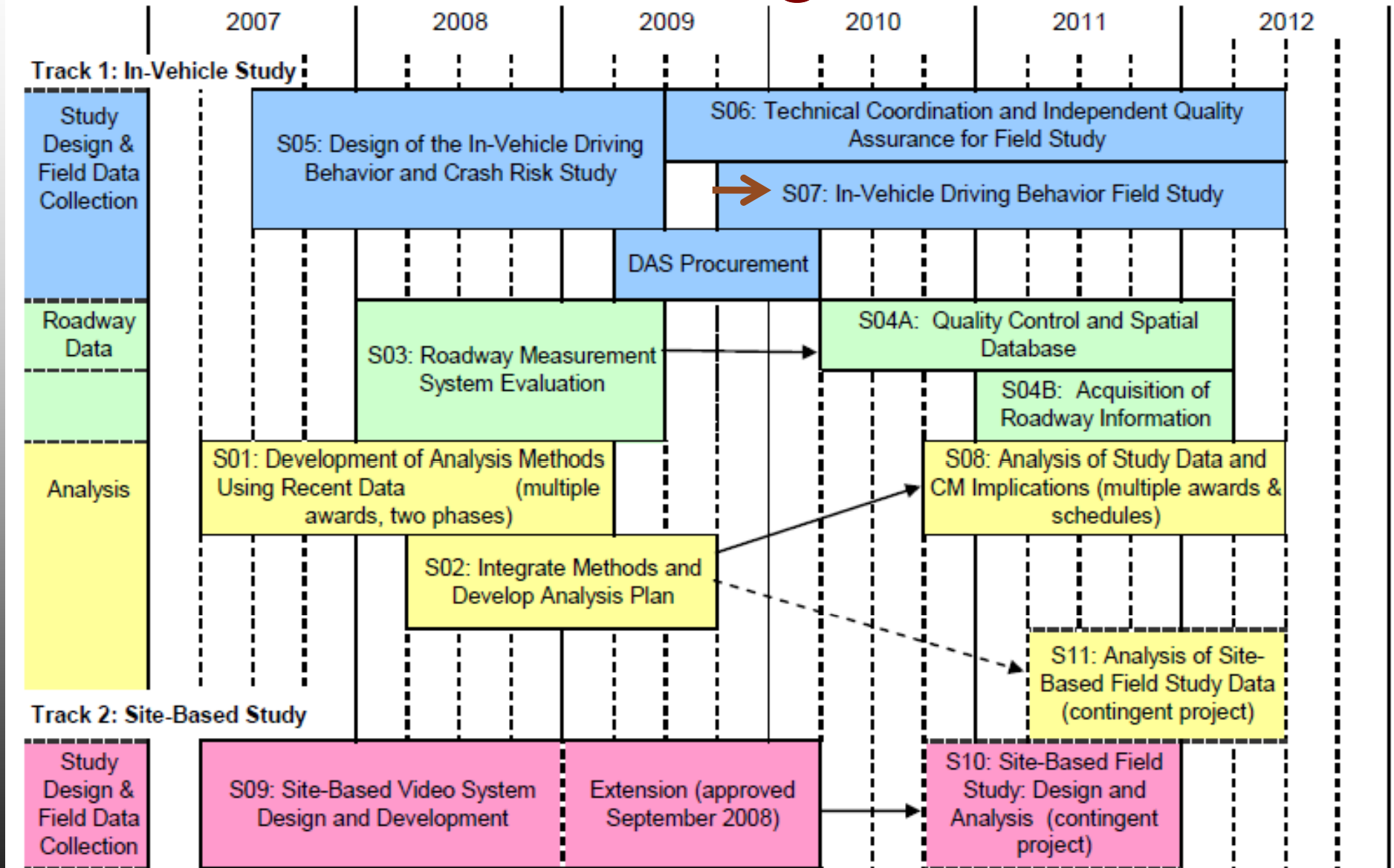
Shane McLaughlin

September 21, 2009

# Discussion Topics

- **Timing**
- **Study Design**
  - Participant and vehicle sample design
  - Data collection sites
  - Participant recruitment
- **Data to be Collected**
  - Driver assessment
  - Crash investigation
  - DAS variables
- **Data Management**
  - IRB Approvals and CoC
  - Collection process: vehicle to server
  - Data processing
  - Expected data volume
  - Quality control
- ***Make Better Use of Naturalistic Data***

# Timing



# Study Design - Summary

- Participant Recruiting Approach Random Listed Sample (centralized)
  - Supplemented Targeted Recruitment (at S07 sites)
- Data Collection Sites 6
  - Total DAS Units 1,950
  - Amount of Data 3900 data-year
  - 150 – 450 DASs per site
- Total Participants ~3,100
  - Duration per Participant 12 or 24 months
- Passenger Car, Van, SUV, Pickup Truck

# Study Design: Objectives

- Data must be usable today and in 20 years
- Data must be as flexible as possible – throw no information away
  - Keep as high a resolution data as possible
- Access must be available for a large range of users with varied capabilities and interests
  - Transportation Researchers
  - State and Federal DOTs
  - OEMs and Tier ones
  - University Researchers
  - Advocates
- Different data and resolution required for different groups

# Participant and Vehicle Design

Gender & Age Range	Age Range Description	One Year	Two Years	DAS units	Primary Participants	Data-Years
M 16-17	Minor Teen	72	28	100	172	200
M 18-20	Adult Teen	72	28	100	172	200
M 21-25	Young Adult	72	28	100	172	200
M 26-35	Adult	72	28	100	172	200
M 36-50	Middle Adult	72	28	100	172	200
M 51-65	Mature Adult	72	28	100	172	200
M 66-75	Younger Older Driver	72	28	100	172	200
M 76+	Older Older Driver	72	28	100	172	200
F 16-17	Minor Teen	72	28	100	172	200
F 18-20	Adult Teen	72	28	100	172	200
F 21-25	Young Adult	72	28	100	172	200
F 26-35	Adult	72	28	100	172	200
F 36-50	Middle Adult	72	28	100	172	200
F 51-65	Mature Adult	72	28	100	172	200
F 66-75	Younger Older Driver	72	28	100	172	200
F 76+	Older Older Driver	72	28	100	172	200
Any	Advanced Vehicle Technology	0	350	350	350	700
<b>Totals:</b>		<b>1,152</b>	<b>798</b>	<b>1,950</b>	<b>3,102</b>	<b>3,900</b>

# Data Collection Sites



# Participant Recruitment

- A central call center managed by single point
- List sample dialing of selected zip codes
- Ideally all participants obtained from the call center
- Traditional means will also be needed to fill cells
  - Likely targeted hard to get groups older drivers, teen drivers, high-end vehicles, etc...

# Participant Recruitment

- Traditional Recruitment
- Common recruitment materials and approaches will be used across sites
  - Newspaper ads
  - Flyers
  - Personal appearances
  - Web advertising (e.g., craigslist.org)
- All phone calls will be done from the central call center. For traditional recruitment they will be given a call in number that reaches the call center
- Recruitment method used will be included in the dataset

# Data to be collected

1. Driver assessment
  2. DAS variables
  3. Crash investigation
-

# Data Acquisition System

- Multiple Videos
  - Machine Vision Eyes Forward Monitor
  - Machine Vision Lane Tracker
  - Machine Vision Driver ID
- Accelerometer Data (3 axis)
- Rate Sensors (3 axis)
- GPS
  - Latitude, Longitude, Elevation, Time, Velocity
- Forward Radar
  - X and Y positions
  - Xdot and Ydot Velocities
- Cell Phone
  - ACN, health checks, location notification
  - Health checks, remote upgrades
- Illuminance sensor
- Passive alcohol sensor
- Incident push button
- Video
- Audio (only on incident push button)
- Turn signals
- Vehicle network data
  - Accelerator
  - Brake pedal activation
  - ABS
  - Gear position
  - Steering wheel angle
  - Speed
  - Horn
  - Seat Belt Information
  - Airbag deployment
  - Many more variables...

# Data Management

- IRB Approvals and Certificate of Confidentiality
- Collection process:  
vehicle to server
- Data processing
- Expected data volume
- Quality control

# IRB Models

- IRB of Record
  - Requires a letter of reliance between the two institutions
  - One institution is relying on another institution for all approval and oversight of human subjects
- Individual IRB approvals
  - Each institution obtains IRB approval for its own activities

# IRB of Record Model

- Advantages
  - Can use consistent consent forms across sites
  - Helps with coordination of necessary amendments
  - Since this will require full board review, having a single board review it will simplify the process
- Technical Coordinator organization (Virginia Tech) will be IRB of Record
- National Academy of Sciences IRB will be involved and provide its own approval

# IRB and Certificate of Confidentiality Approach

- Currently working among IRBs
- Draft IRB of Record provided
  - Discuss all changes that will be required to use IRB of Record
  - Probably require multiple post-meeting iterations
  - Likely that some will still require their own IRB but it must be compatible with IRB of Record and the SHRP2 data collection

# IRB and Certificate of Confidentiality Approach

- Once IRB of Record has incorporated the necessary changes, submit to Virginia Tech's IRB
- As necessary collection sites submit IRB
- Once all/majority of IRBs are approved, submit a Certificate of Confidentiality to governmental agency for human subjects (National Institute of Health)
- Finally obtain National Academy of Sciences IRB approval
  - It would be helpful if this could be done in parallel with certificate of confidentiality

# Data Summary

- All raw data (video and sensor/parametric) will be recorded continuously and will be preserved.
- Data will be encrypted at the time of acquisition on the DAS, and the data will not be decrypted until the data reach the VTTI servers.
  - DAS can be placed in diagnostic mode for quality checks of instrumentation
- Data will be downloaded from each DAS at ~4 month intervals.
  - Health Check includes remaining hard drive capacity
- Each collection site will have a data sever for downloading, transferring data and refurbishing hard drive

# Data Processes:

## Remove Proprietary Information

- The data will be decrypted prior to loading in the database or on a file server in the case of video files
- OEM proprietary or sensitive information will be removed.

---

  - Depending on the OEM different information will be available
  - Working with each OEM, a strategy will be developed to determine what information is sensitive and what can be included.
    - For example, a bit vector may carry a proprietary state indication along with one for general use.
  - Some processing will be needed to desensitize/aggregate the data so as to not communicate proprietary information.

# Summary Storage

- Capacity for storing specific data streams at different collection rates is necessary (asynchronous)
- Data stamped with real-time clock
- Excluding additional supplemental databases and synthesized data, there is an estimated 800TB of video and 100TB of raw sensor data expected from this study.
- A peta-scale data repository will be required to house these data with appropriate Redundant Array of Independent Disks (RAID) configuration

# Mission Control Software

- Mission Control Software being developed
  - Allows tracking vehicles, participants, and DASs across sites
  - Participant recruitment facilitation/coordination
  - Manage Automated Collision Notification
  - Manage health checks, etc...

# Health Check

- Health Check information is sent to mission control software on a regular basis from DAS
  - Health checks include sensor function, video, simple relationships between variables.
  - Health check includes DAS Hard Drive capacity measure
- recommends when vehicle needs to be visited for health check or data download – consistent across sites
- Collection site can sign into software and review their health checks
- Collection site fits it into their schedule for maintenance and downloads
  - Participant is contacted and a time and location is determined for the maintenance call

# Data Access and Support

- A role based user security will be used to grant researchers access to the data.
- No specialized proprietary software needed. The data can be analyzed with traditional commercial software such as SAS, R, and Matlab.
  - A visualization tool that will allow users to view data and video together using Matlab with open source code will be developed
- Data dictionaries have been developed that will help users understand both what the data is and how it was coded.
- Currently providing event based data from 100Car study through website.

---

**Thank You**