Statistical Models for Combining Information: Stryker Reliability Case Study

Rebecca Dickinson, Virginia Tech

Laura Freeman, IDA

Alyson Wilson, IDA

Bruce Simpson, IDA

IDA

Outline

- **The Stryker Family of Vehicles**
- **Motivation for Using All Information**
- **Methods**
  - Exponential versus Weibull Distribution
  - Frequentist versus Bayesian Methodologies
- **Results**
- **Conclusions**

Bottom Line Up Front

- **The purpose of this case study is to illustrate proof of concept**
  - Stryker OT dataset is robust
  - Common chassis, multiple variants
- **Support integrated testing**
  - How do we leverage all data in quantitative statistical analyses?
- **Results:**
  - Tighter confidence intervals
  - Better reliability estimates
  - Benefits are greatest for vehicles with only 0-2 reported failures in OT
- **Future Directions**
  - Stryker case study shows value-added
  - How do we use this in future analyses?
  - How do we use this in scoping future test plans?

The Stryker Family of Vehicles

Infantry Carrier Vehicle

Engineer Squad Vehicle

Mortar Carrier Vehicle

Stryker System Description

The Stryker family of vehicles includes 10 separate systems.

- **Two Basic Vehicle Variants**
  1. Infantry Carrier Vehicle (ICV) - the infantry/mission-vehicle type
     - Base vehicle for nine separate configurations
       - Infantry Carrier Vehicle (ICV)
       - Mortar Carrier Vehicle (MCV)
       - Antitank Guided Missile Vehicle (ATGMV)
       - Reconnaissance Vehicle (RV)
       - Fire Support Vehicle (FSV)
       - Engineer Squad Vehicle (ESV)
       - Commander's Vehicle (CV)
       - Medical Evacuation Vehicle (MEV)
       - NBC Reconnaissance Vehicle (NBCRV)*

     (ICV through MEV: Considered in this analysis)
  2. Mobile Gun System (MGS)* – direct fire platform and performs the maneuver fire support role.

**NBCRV and MGS were not included because they were on a different acquisition timeline*

Stryker Mission Essential Functions

- **There are four essential functions**
  - Move
  - Shoot
  - Command and Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR)
  - Survive
- **A failure is an event in which an item or part of an item does not perform as specified.**
- **The Army failure definition scoring criteria (FDSC) categorizes the severity of failures**
  - **System Abort**
    - » The vehicle is unable to complete the mission.
  - Essential Function Failure
  - Non-essential Function Failure
- **Reliability requirement:**
  - 1,000 mean miles between **system aborts**

Developmental and Operational Testing

Developmental Testing

- **Controlled Conditions**
- **Experienced Technicians operating the vehicles.**
  - They have done this for years and they know the courses really well.
- **Courses**
  - Use courses that are designed to replicate the primary roads, secondary roads, and trail like conditions.

DT And OT Are Different!

- Operators
- Environments
- Test Durations

Operational Testing

- **Operational Conditions**
- **An army unit comes in to do this testing.**

- **Courses**
  - OT data set comes from testing that was done at Fort Knox
  - Most of the testing was done using secondary road type conditions.
- **Limited amount of Time**
  - Due to operator availability and range availability
  - Operational testing may be too short to discover many reliability deficiencies

Motivation For Using All Information

- **What is the Current Practice?**
  - DOT&E in most cases uses only operational test data for reliability analyses
    - » Stryker Beyond Low Rate Initial Production (BLRIP) Report
    - » Benefit: ensures data is representative of operational test conditions
    - » Drawback: discards information from previous testing that provides information on system reliability
- **Why use all test data?**
  - Testing is expensive
  - Lose valuable information by not using all information
- **National Research Council Studies**
  - *Statistics, Testing and Defense Acquisition*, **1998**
    - » Emphasizes that all relevant information be examined for possible use in both the design and evaluation of operational tests …
    - » State-of-the-art statistical methods for combining information should be used, when appropriate, to make tests and their associated evaluations as cost-efficient as possible.
  - *Improved Operational Testing and Evaluation*, **2006**
    - » Focuses specifically on methods of combining information for the Stryker family of vehicles.

Keys to Success

- **Eliminate or account for as many sources of variation as possible**
  - Common response variable across test phases:
    - » Reliability data
      - Consistent data collection and scoring
      - Detailed data records including:
        - Miles between each abort (not just total miles and total aborts)
        - Sub-system records for each abort
- **Leverage all common information**
  - Family of Vehicles: allows us to pool information by leveraging relationships between vehicles
- **Think hard about the model!**

Reliability Analysis

- **Reliability is an essential component operational suitability**
- **Examples of reliability data:**
  - Miles driven until failure, hours of use until a failure, number of on-off cycles until a failure
- **Commonly used reliability distributions:**

Exponential Distribution

- Historically used in DoD reliability assessment
- Simple model: only one parameter to estimate

$$f(t_i) = \frac{1}{\lambda} e^{-\left(\frac{t_i}{\lambda}\right)}$$

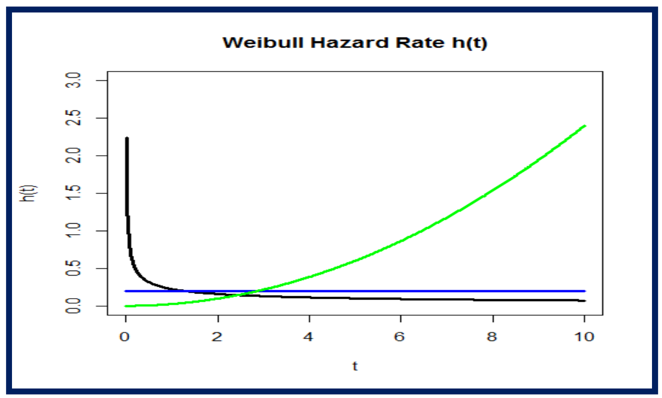
- Easy to interpret: under this parameterization, λ is the mean time between failures

Weibull Distribution

- Flexible distribution: two parameters

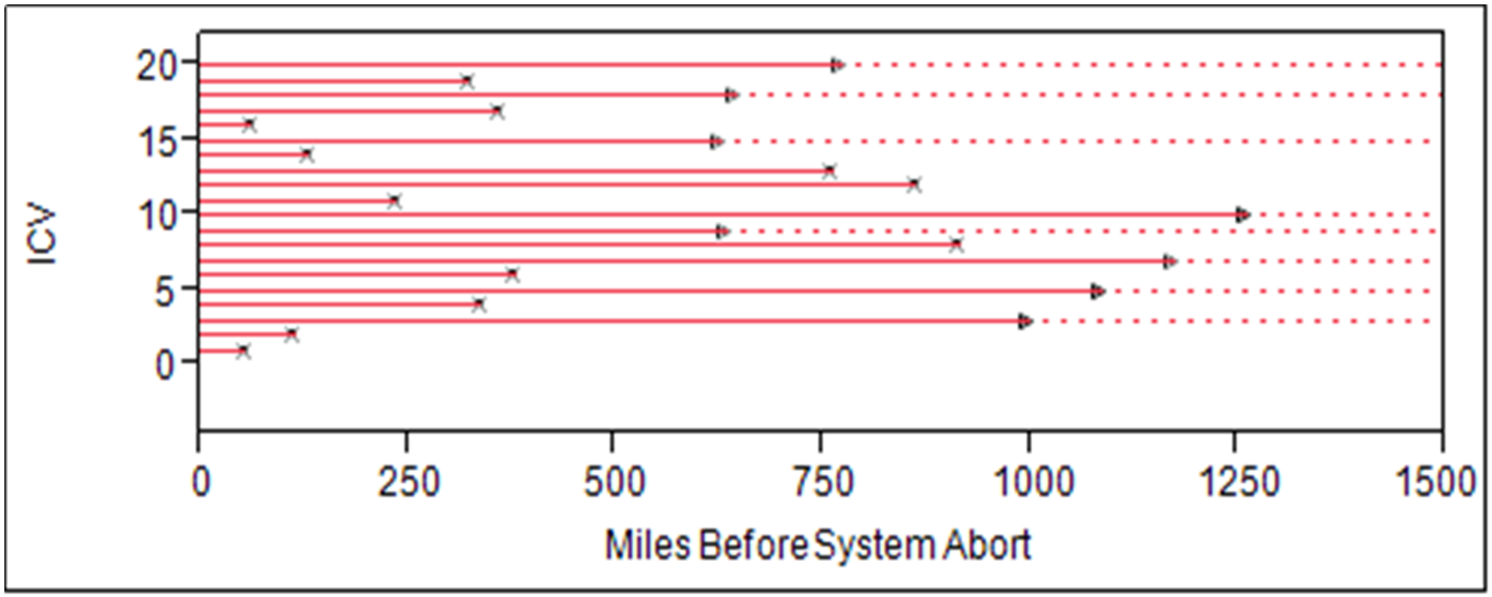
$$f(t_i) = \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1} e^{-\left(\frac{t_i}{\eta}\right)^{\beta}}$$

- Can describe multiple failure mechanisms



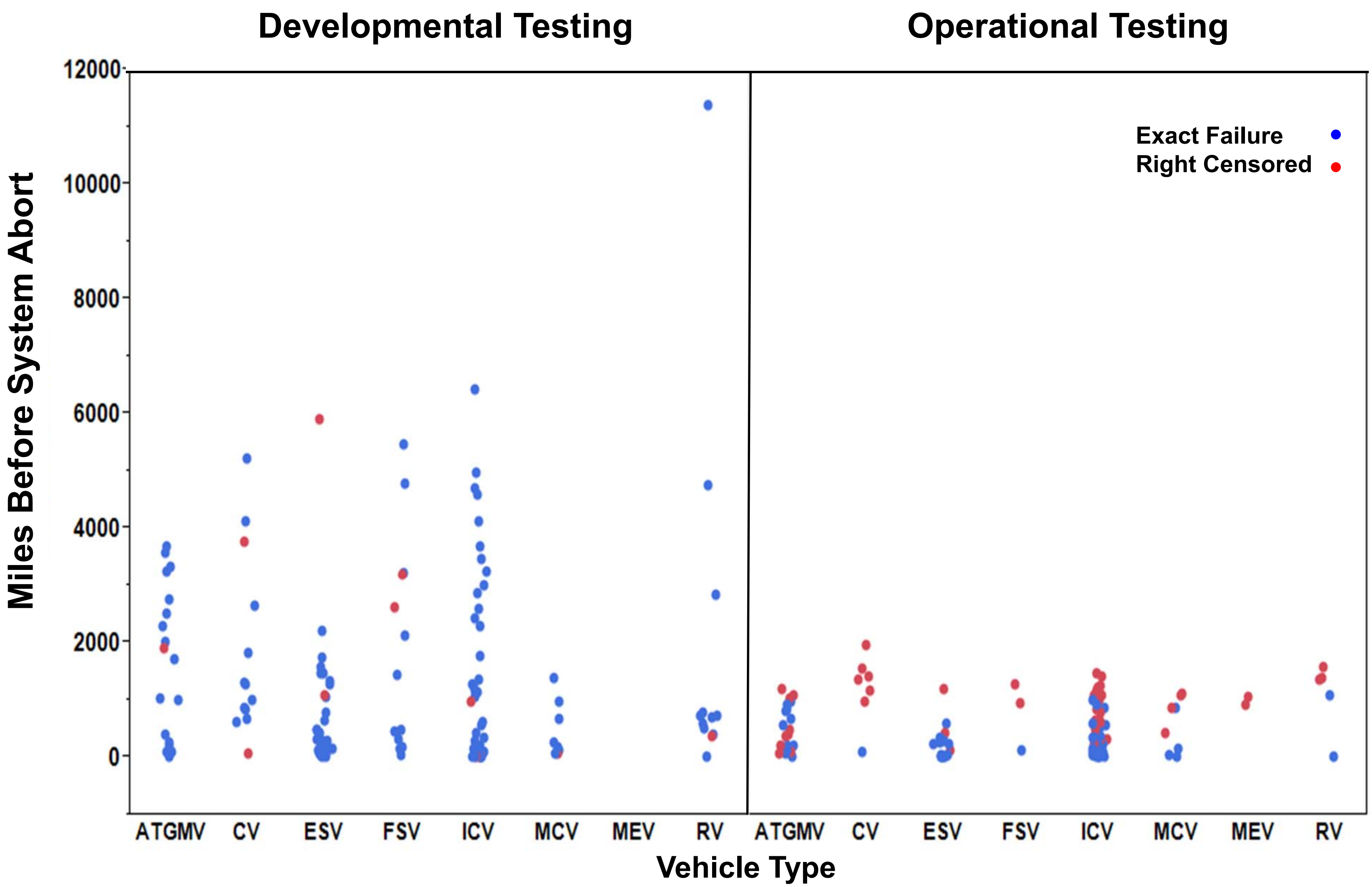
Unique Features of Reliability Data

- **The exact failure times are not always known.**
  - When this happens we say that the data is censored.



- **No negative data values (failure times > 0).**
  - We model reliability data using distributions for positive random variables.
  - Popular choices of distributions to use include the exponential and Weibull.

The Stryker 2003 Data Set



A Traditional Analysis - Using OT Data Only

The chart below is similar to that which was included in the report written for DOT&E when considering this data set.

We will be using this as our reference when comparing the new methods that combine information across the developmental and operational test phases.

Stryker Reliability by Variant using Operational Test Data

Vehicle Variant	Total Miles Driven	System Aborts	MMBSA	MMBSA 95% LCL	MMBSA 95% UCL
ATGMV	10334	12	861	492.9971	1666.62
CV	8494	1	8494	1524.505	335495.1
ESV	3771	13	290	169.6326	544.7885
FSV	2306	1	2306	413.8815	91082.13
ICV	29982	35	857	615.9437	1229.84
MCV	4521	4	1130	441.4354	4148.219
MEV	1967	0	-	656.6007	-
RV	5374	2	2687	743.8384	22187.42
Total	**66749**	**68**	**982**	774.2946	1264.074

$$\text{Mean Miles Before a System Abort } (\mathbf{MMBSA}) = \frac{\text{Total Miles Driven}}{\text{System Aborts}}$$

IDA

Failure-Time Regression Models

We began by using the exponential distribution to model the miles before a system abort

$$t_{ijk} \sim exponential(\lambda_{ij})$$

$i = 1,2$ (test phase)
$j = 1,2, \ldots, 7$ (vehicle variant)
$k = 1,2, \ldots, n_{ij}$ (miles)

We can express rate parameter, λ, as a function of explanatory variables to find an estimate for MMBSA

Model 1:

Average over vehic… …mes vehicle type does not matter)

$$\lambda_{i.} = \gamma_0 + \gamma_1 \text{Test Phase}$$

Naïve : we know variant and test phase impact reliability

Model 2:

Average over test phase (assumes test phase does not matter)

Yes, we combine information – but we completely ignore the test phase!

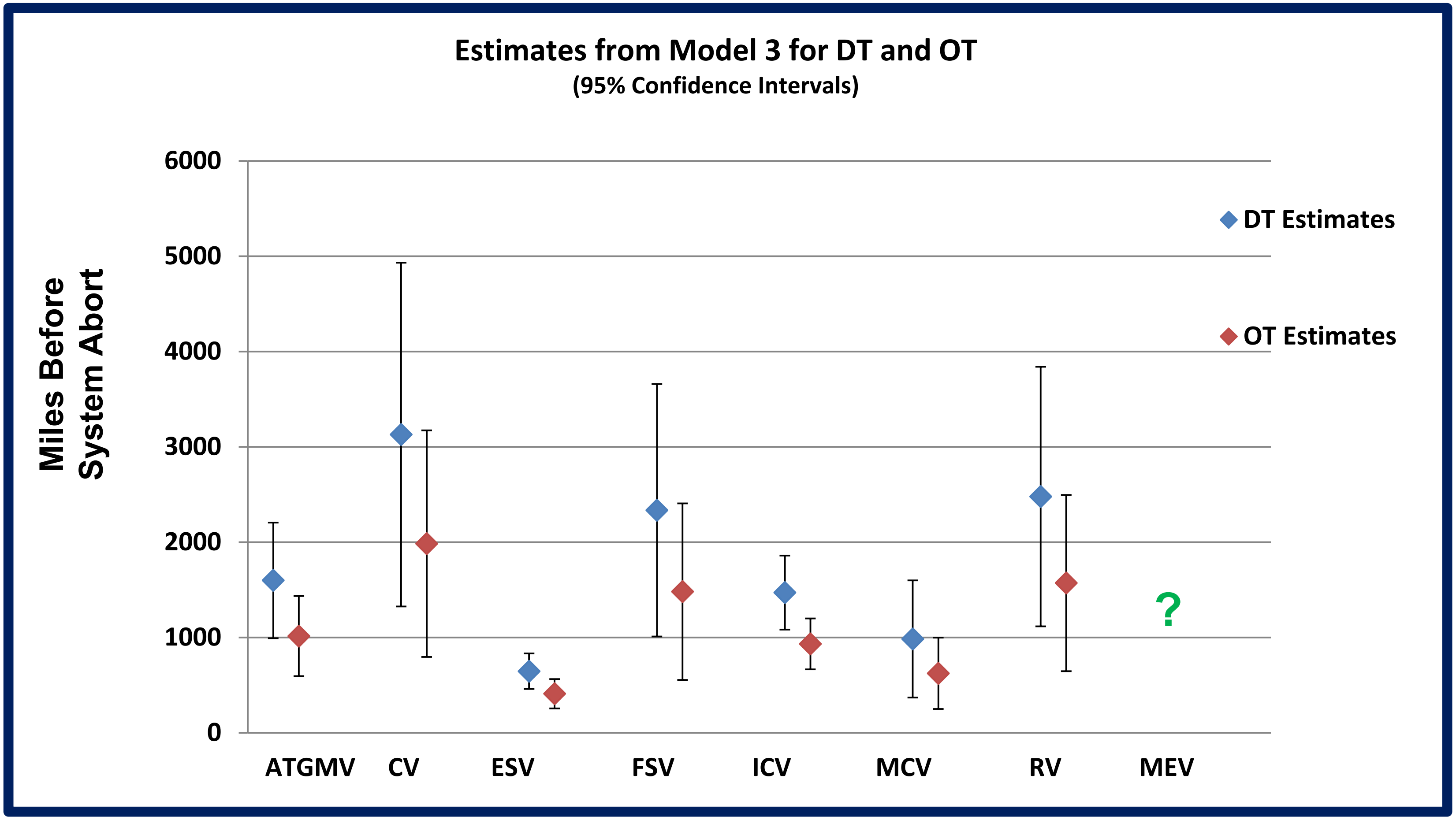
$$\lambda_{.j} = \gamma_0 + \gamma_1 \text{ATGMV} + \ \ldots + \gamma_6 \text{MCV}$$

Model 3:

Look at differences based on Test Phase & Vehicle Type.

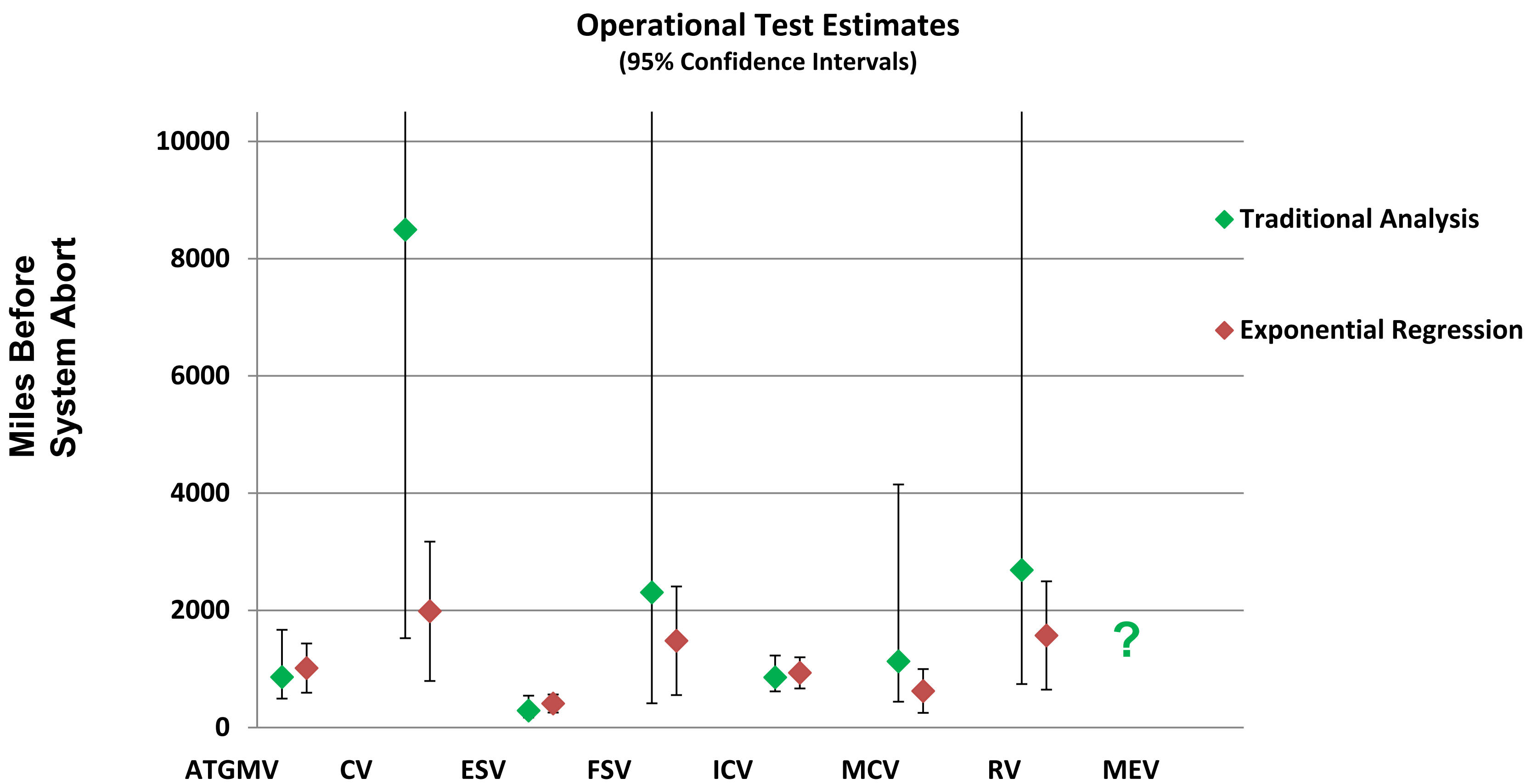
$$\lambda_{ij} = \gamma_0 + \gamma_1 \text{Test Phase} + \gamma_2 \text{ATGMV} + \cdots + \gamma_7 \text{MCV}$$

Exponential Regression Results



This model estimates a 37% reduction in the MMBSA moving from DT To OT

Comparing Confidence Intervals



Tighter confidence intervals & better estimates for MMBSA

Bayesian Analysis

- **Bayesian models still require a parametric statistical model**
  - Bayesian model is specified by:
    - » Parametric statistical model (just as before)
    - » Prior distribution
  - Bayes Theorem: posterior distribution is proportional to the likelihood (data) times the prior

- **Why might we want to consider this option?**
  - Incorporate more information through the use of a prior
    - » A degradation from DT to OT
    - » This allows for us to come up with an estimate for the Medical Evacuation Vehicle (0 observations in DT and 2 censored observations in OT) by using the information that we know about the other vehicles.
  - Ease of inference

We can incorporate more information!

Bayesian Models Considered

Bayesian Model 1

$$\boldsymbol{t_{DT} \sim exp(\lambda) \quad t_{OT} \sim exp(\lambda/\eta)}$$

Using Non-Informative Priors:

$$\lambda \sim gamma(.001, .001)$$
$$\eta \sim beta(1,1)$$

Comparable to the Failure-time Regression Model 1

Bayesian Model 2

$$\boldsymbol{t_{DT} \sim exp(\lambda_i) \quad t_{OT} \sim exp(\lambda_i/\eta)}$$

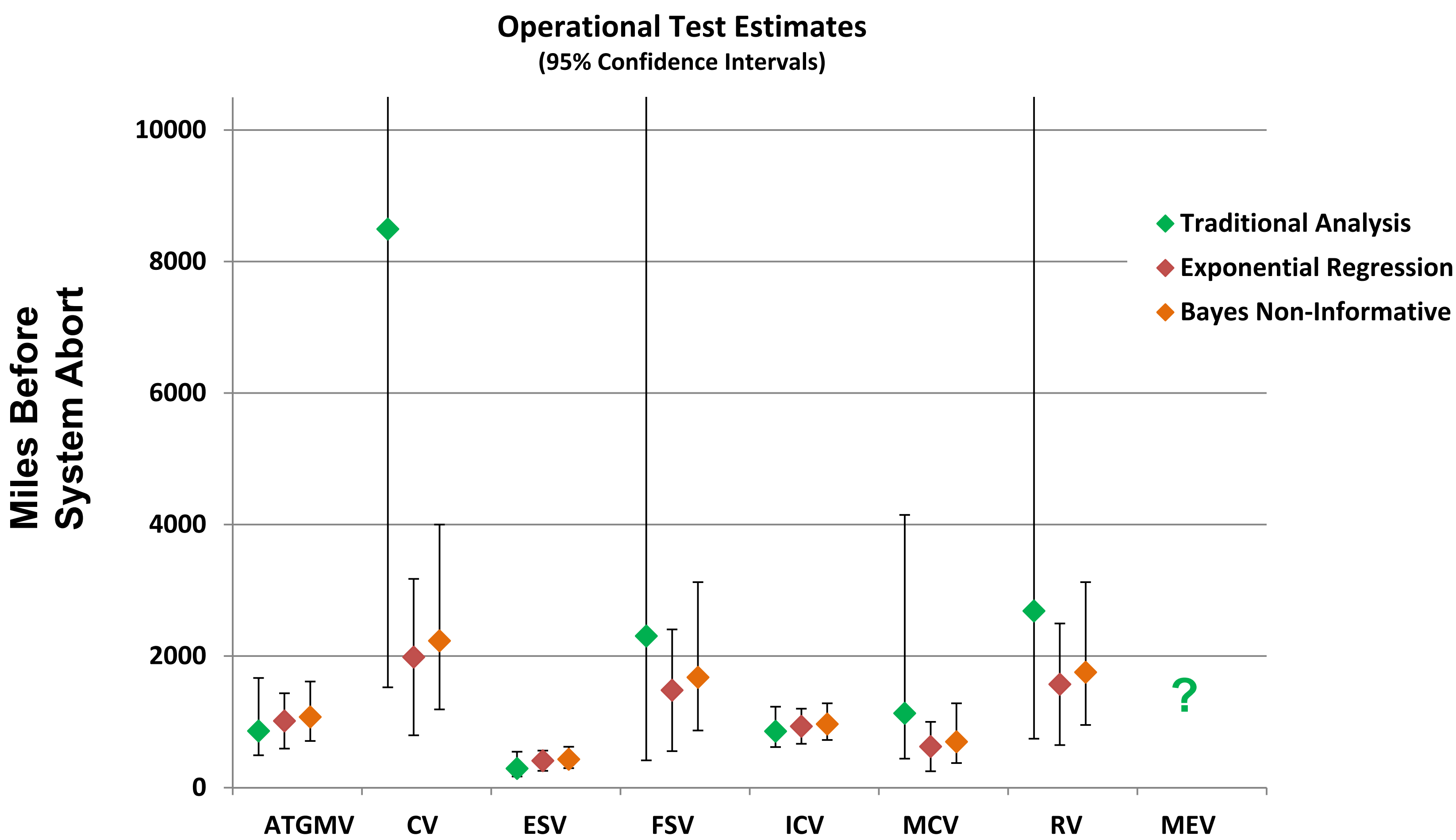
$i = 1,2, \dots, 7$ (vehicle variants)

Using the Non-Informative Priors:

$$\lambda_i \sim gamma(.001, .001)$$
$$\eta \sim beta(1,1)$$

Comparable to the Failure-time Regression Model 3

Comparing Intervals



Point and interval estimates for MMBSA are nearly identical

Incorporating More Information

- **Informative Priors**
  - Based on subject matter expertise
    - » Data is already included in model
- **Hierarchical Models**
  - Assumes the parameters are related, the data tells us how closely related
  - Hierarchical models for the Stryker case study allow us to estimate MEV reliability based on other data

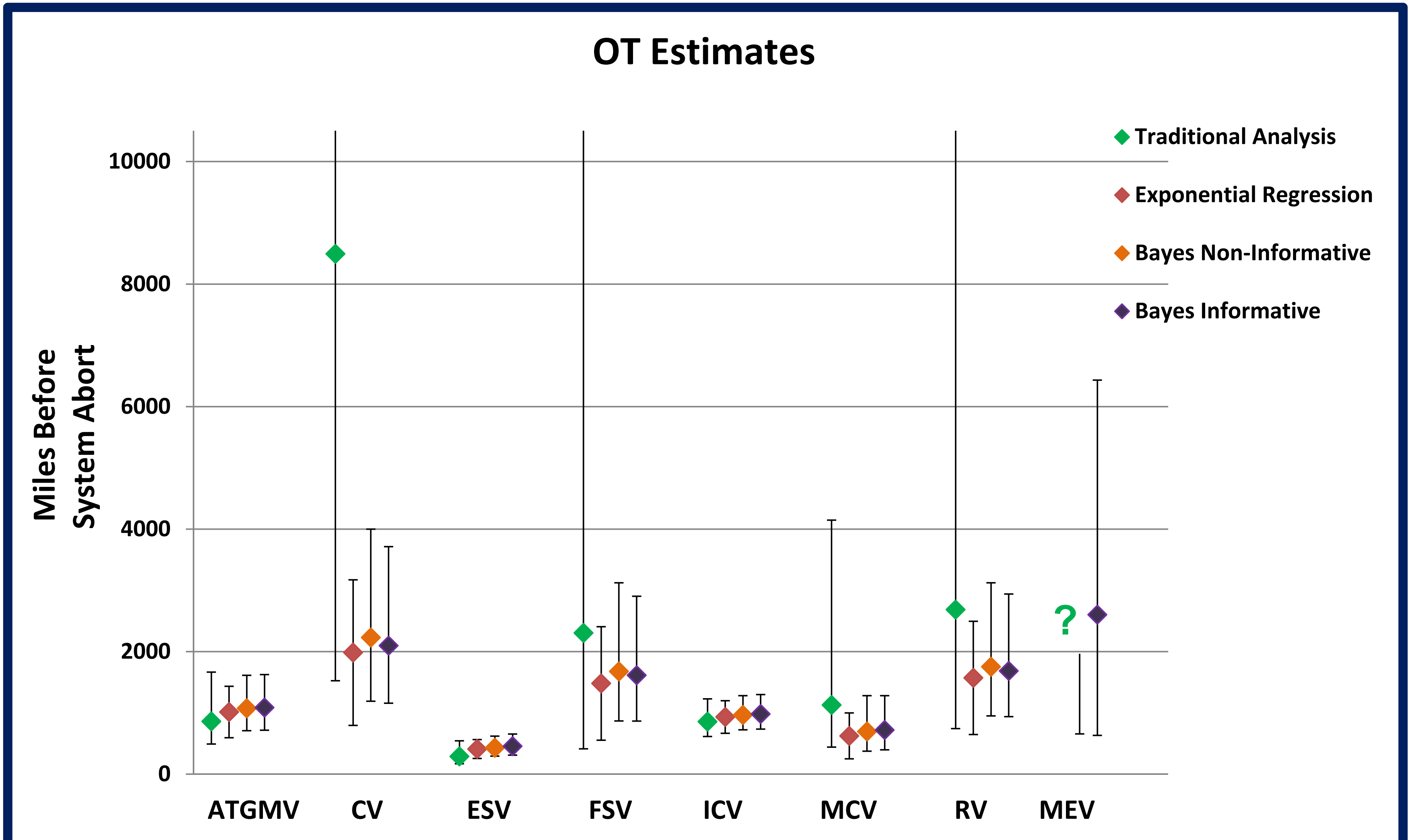
A Model That Allows Us To Estimate MEV Reliability

$$t_{DT} \sim exp(\lambda_i) \quad t_{OT} \sim exp(\lambda_i/\eta)$$

$i = 1,2,\ldots,8$ (vehicle variants including MEV)

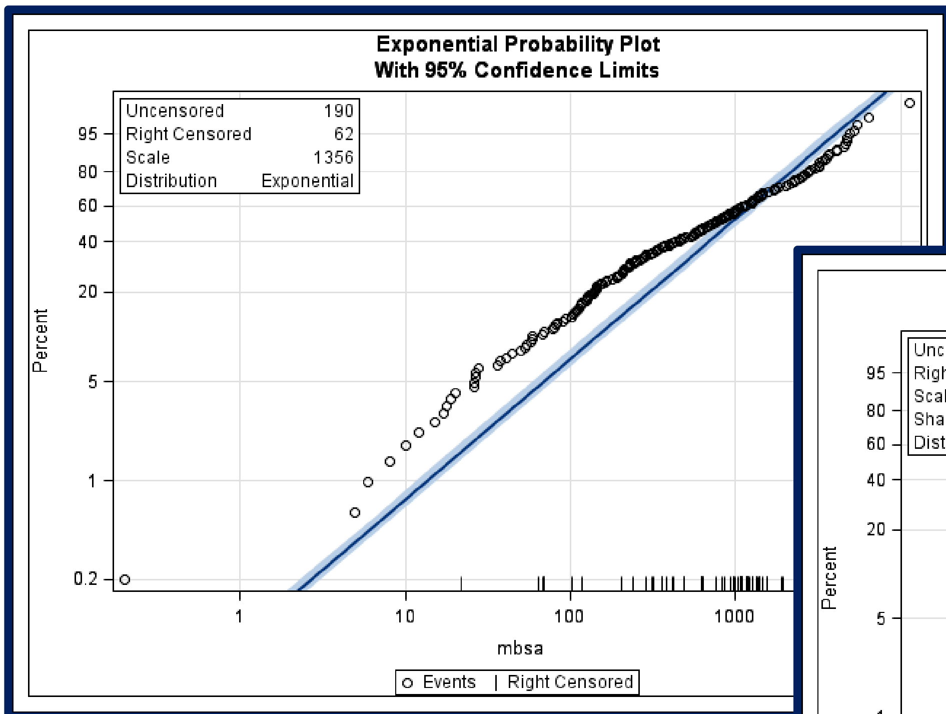
$$\lambda_i \sim gamma(a,b)$$
$$\eta \sim beta(1,1)$$
$$a \sim gamma(.001,.001)$$
$$b \sim gamma(.001,.001)$$

Results: Incorporating More Information

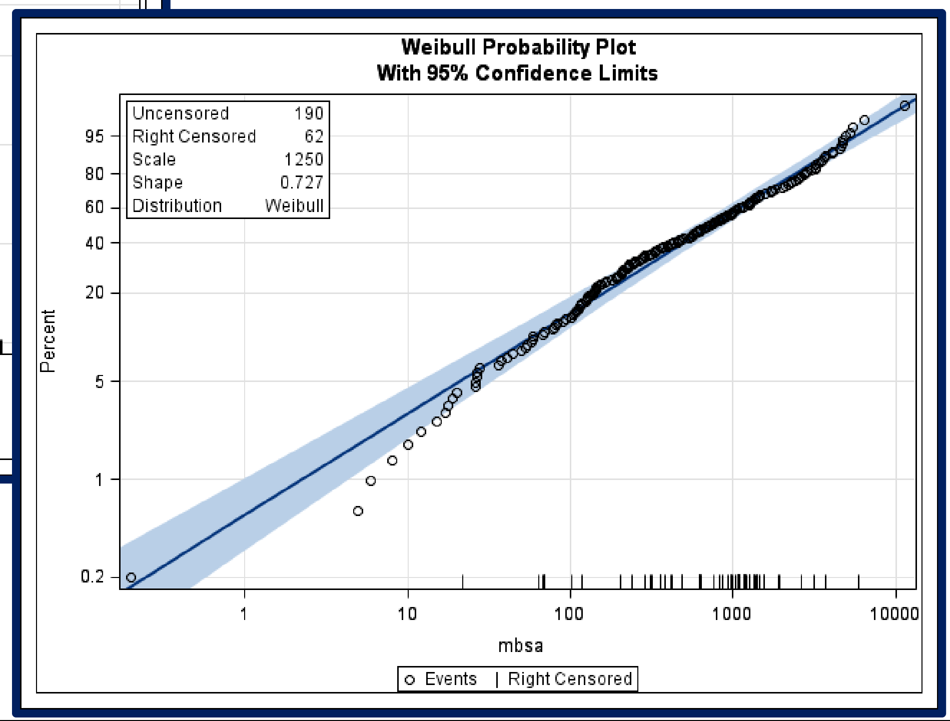


Reality Check

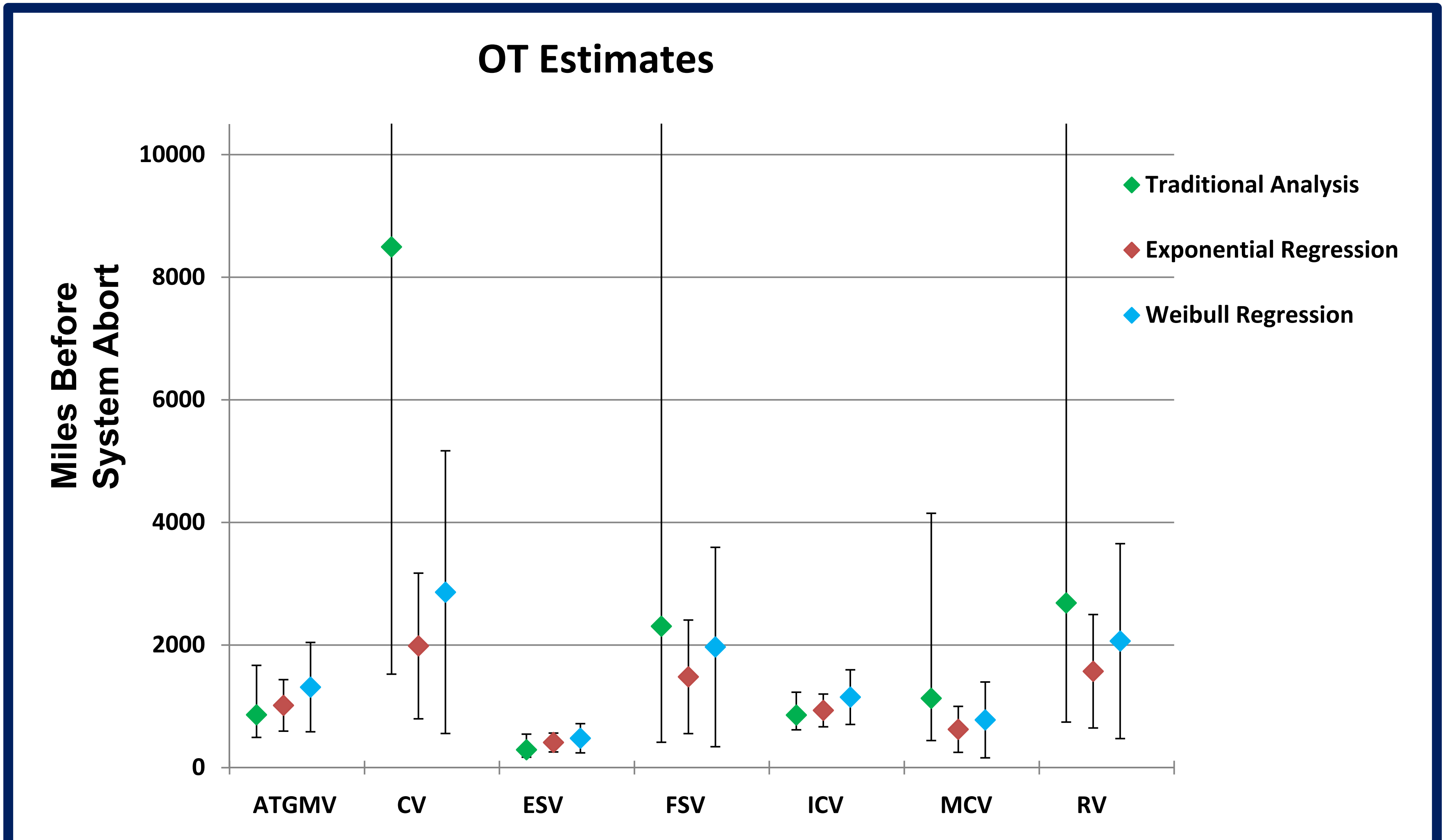
- **Is the exponential distribution appropriate?**



  – Weibull Distribution is more flexible
  – Weibull Distribution fits the data better



Comparison of Exponential and Weibull Results



IDA

Statistical Challenges

- **Likelihood based inferences**
  - Cannot always be done in standard statistical software
  - Multivariate delta method
- **Censored data**
- **Might need to write your own code.**
  - Software packages don’t always provide enough flexibility
- **No data set is ever perfect**
  - Missing data
    - » Multiple imputation
    - » Bayesian imputation

Model Selection Considerations

- **Ease of use**
  - Exponential regression available in JMP
  - Bayesian techniques require code writing
  - Explanation of results
- **Frequentist versus Bayesian**
  - Interpreting confidence intervals (credible intervals)
  - Zero failures – point estimates only exist in a Bayesian framework
  - Can we incorporate information from data directly?
    - » Bayesian models allow us to incorporate information only available as summary statistics
- **Informative versus Non-informative priors**
  - Is there reliable subject matter expert information to incorporate?

Conclusions

- **We can use basic statistical models to incorporate information from multiple testing phases into OT assessments.**
- **The results are:**
  - Tighter confidence intervals (an average of a **60%** reduction in the interval width)
  - Better estimates for MMBSA
    - » Commander's Vehicle estimates were optimistically high before incorporating information from DT
  - Benefits are greatest for vehicles with only 0-2 reported failures in OT
- **Model specification requires careful consideration**
  - If the model is wrong the results are not meaningful
- **Bayesian techniques provide:**
  - Ability to incorporate more information than is contained in the data
    - » Subject matter expertise
    - » Historical information not directly contained in data
  - Ease of inference
    - » Missing data imputation
    - » Censored data with complex likelihoods
- **Analysis requires more statistical knowledge than the Traditional OT analyses**
  - Information gained is worth the effort

IDA

Caveats and Future Directions

- **Concerns**
  - Need operational, statistical, and system engineering expertise to make this work
  - Model specification is key, the model must be appropriate for the data
  - Analyses are nontrivial compared to current standard analyses

- **Future Directions**
  - How do we use this in future analyses?
  - How do we use this in scoping future test plans?

Backup Slides

Summarizing Confidence Intervals

Vehicle	Reduction in Intervals (compared to Traditional Analysis) Under the Assumption t ~ Exponential
ATGMV	0.25
CV	0.99
ESV	0.13
FSV	0.98
ICV	0.10
MCV	0.77
RV	0.91
MEV	
Column Average	0.59

Weibull Model Specification

- **Weibull distribution has two parameters, β and η**
  - Both could be impacted by test phase (DT/OT) and vehicle variant
  - Considered two models:
    - » Both β and η as a function of variant and test phase
    - » Only η as a function of variant and test phase
  - Test phase did not impact the model shape parameter, β

$$f(t_i) = \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1} e^{-\left(\frac{t_i}{\eta}\right)^{\beta}}$$

Reliability and Reliability Data

- **Reliability is an essential component of the assessment of operational suitability of major defense systems.**

- **We can think of reliability as quality over time.**

  One comes to expect that a system, vehicle, machine, or device will perform its intended function under its appropriate operating conditions for some specified period of time.

- **We use data to help predict and assess various aspects of product reliability.**

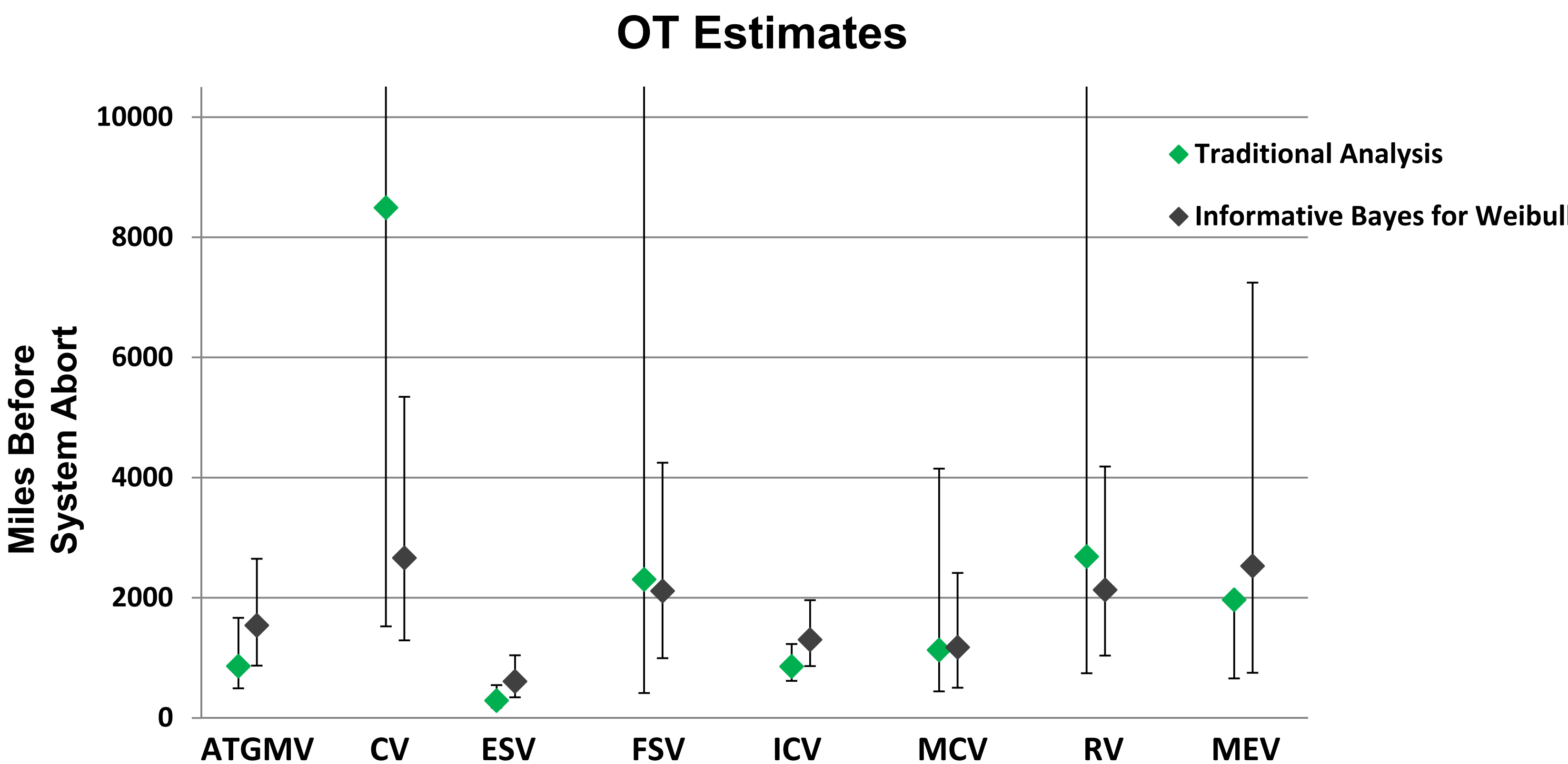
- **Some examples of reliability data include:**

  Miles driven until failure, hours of use until a failure, number of on-off cycles until a failure, …

Failures Are What We Care About

- **We can use Bayesian methods here as well**
  - The provide ease of computation in this case, multiple imputation versus Bayesian imputation (more on next slide)

Results: Incorporating More Information



We can get an estimate for MEV by incorporating the information that we know about the other vehicles.