


# Institutionalizing a Culture of Statistical Thinking in DoD Testing

Dr. Catherine Warner  
Science Advisor



Statistical Engineering Leadership Webinar  
25 September 2017

# Outline

- 
- Overview of DoD Testing
  - Improving Operational Testing
  - Statistical Analysis Methods for Improving Mission Characterization
  - Continuing the Path Forward
    - Bayesian Methods for Maximizing Information
    - Defensible Surveys – Capturing Human Interactions
    - Improving Modeling and Simulation
  - Looking to the Future

# Goal of Operational Test: Evaluate Operational Effectiveness and Suitability

Operational Environment



Representative Users

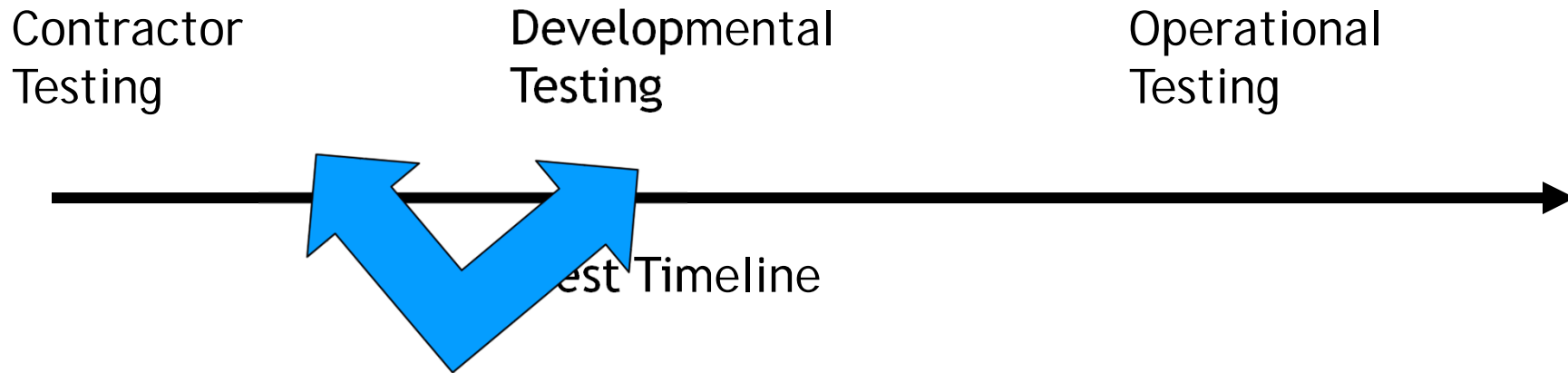


“Real” Threats

Conducting Missions



# DoD Test Paradigm



Tend to be  
requirements driven

Requirements documents are often missing important mission considerations



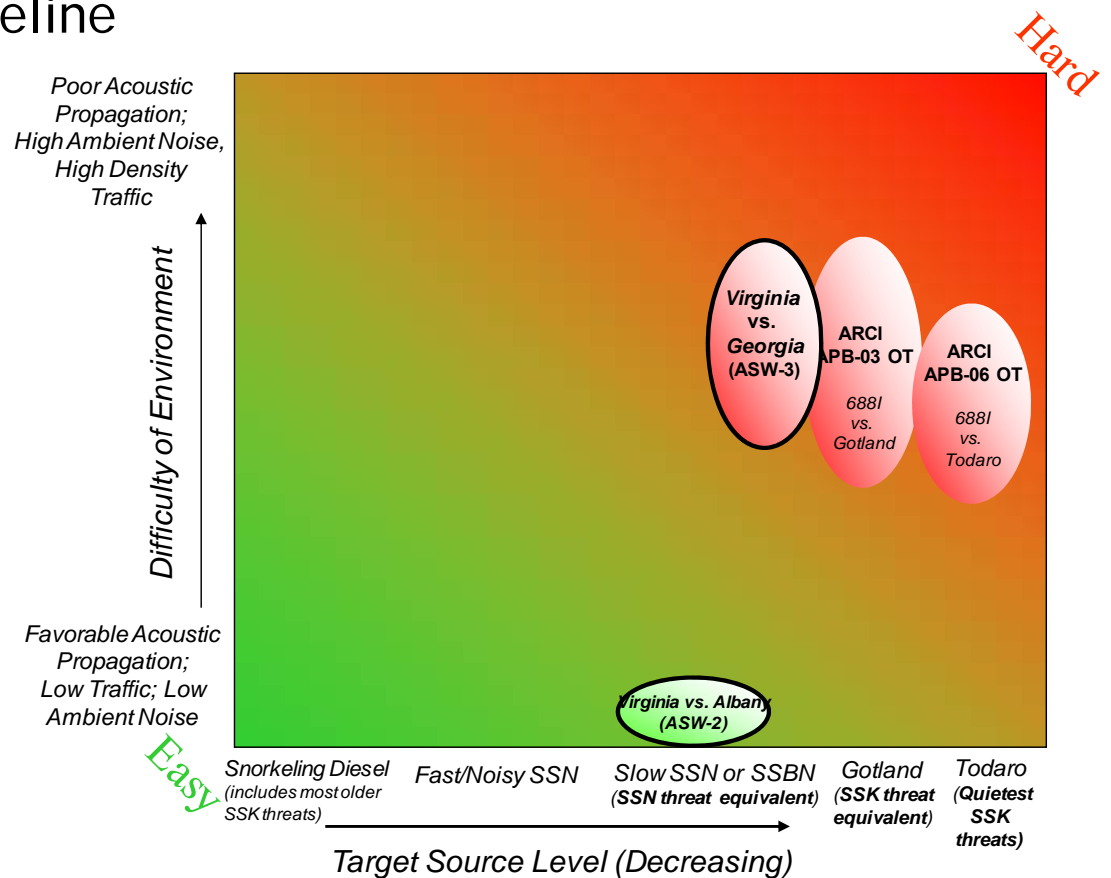
# OT characterizes mission capability

Contractor  
Testing

Developmental  
Testing

Operational  
Testing

## Test Timeline



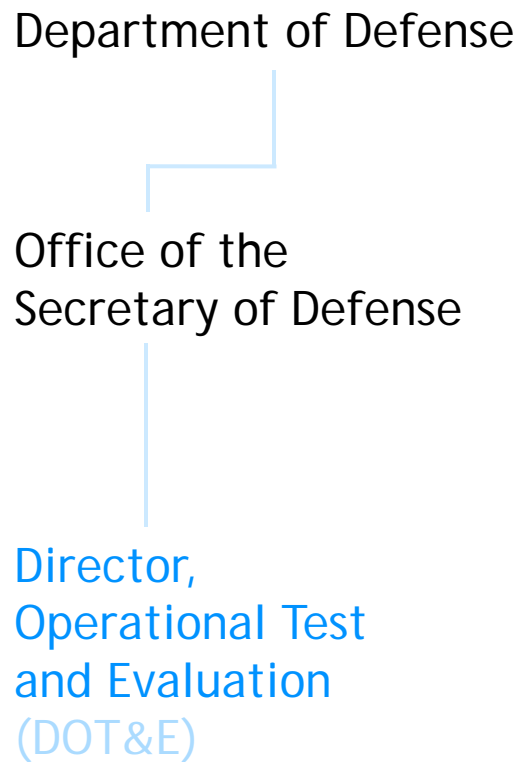


By the early 1980s,  
Congress' concerns were growing



# Congress established DOT&E separate from the Services' operational testing agencies

Department of Defense



```
graph TD; A[Department of Defense] --> B[Office of the Secretary of Defense]; B --> C[Director, Operational Test and Evaluation (DOT&E)];
```

Office of the  
Secretary of Defense

Director,  
Operational Test  
and Evaluation  
(DOT&E)



# Congress established DOT&E separate from the Services' operational testing agencies

Department of Defense

Office of the  
Secretary of Defense

Director,  
Operational Test  
and Evaluation  
(DOT&E)

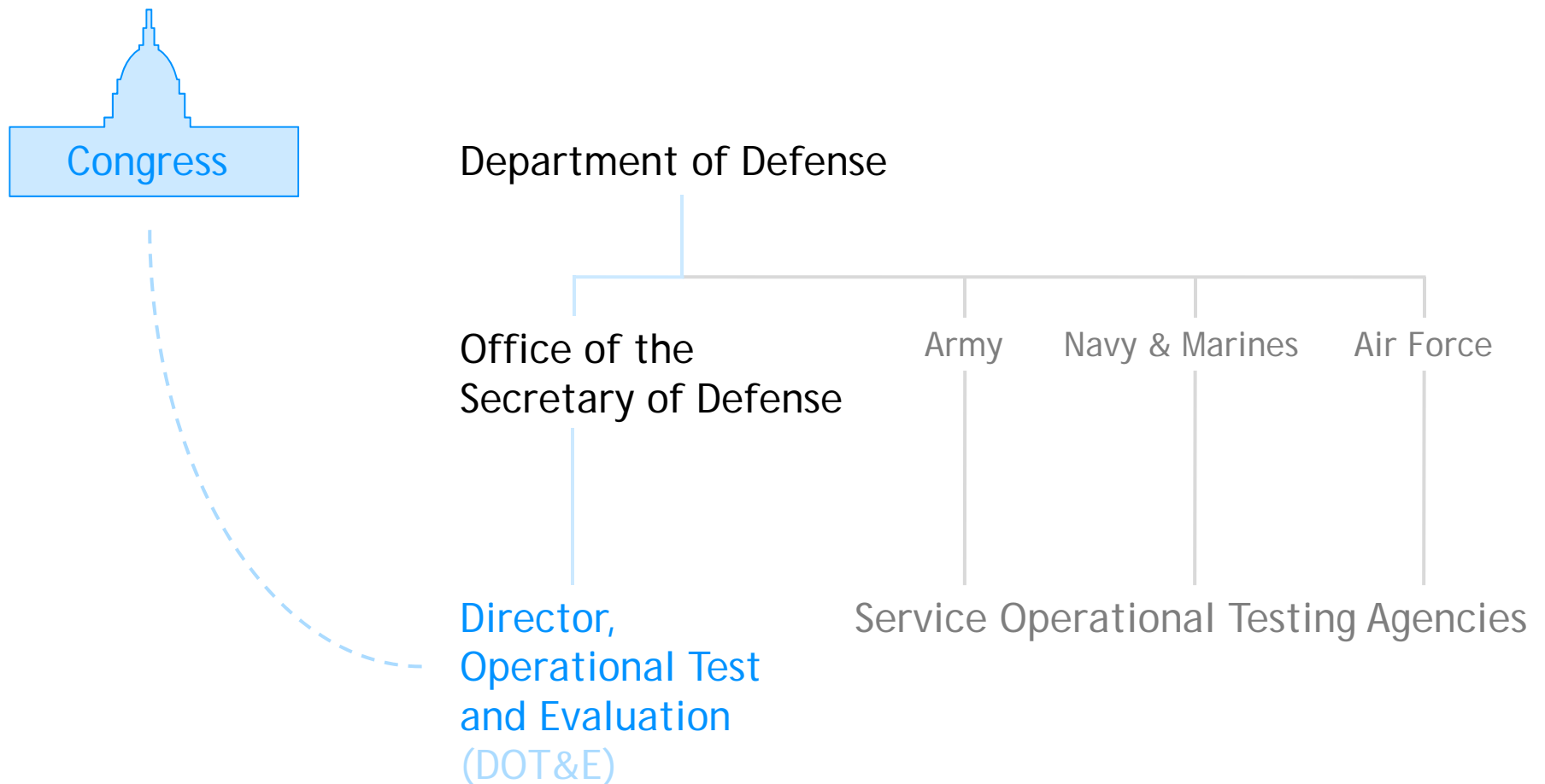
Army

Navy & Marines

Air Force

Service Operational Testing Agencies

# Congress established DOT&E separate from the Services' operational testing agencies



Operational testing provides critical information to warfighters about new systems...

*Before warfighters' lives and missions depend on them*



Operational testing provide critical information to warfighters about new systems...

*Before warfighters' lives and missions depend on them*

Time to correct problems

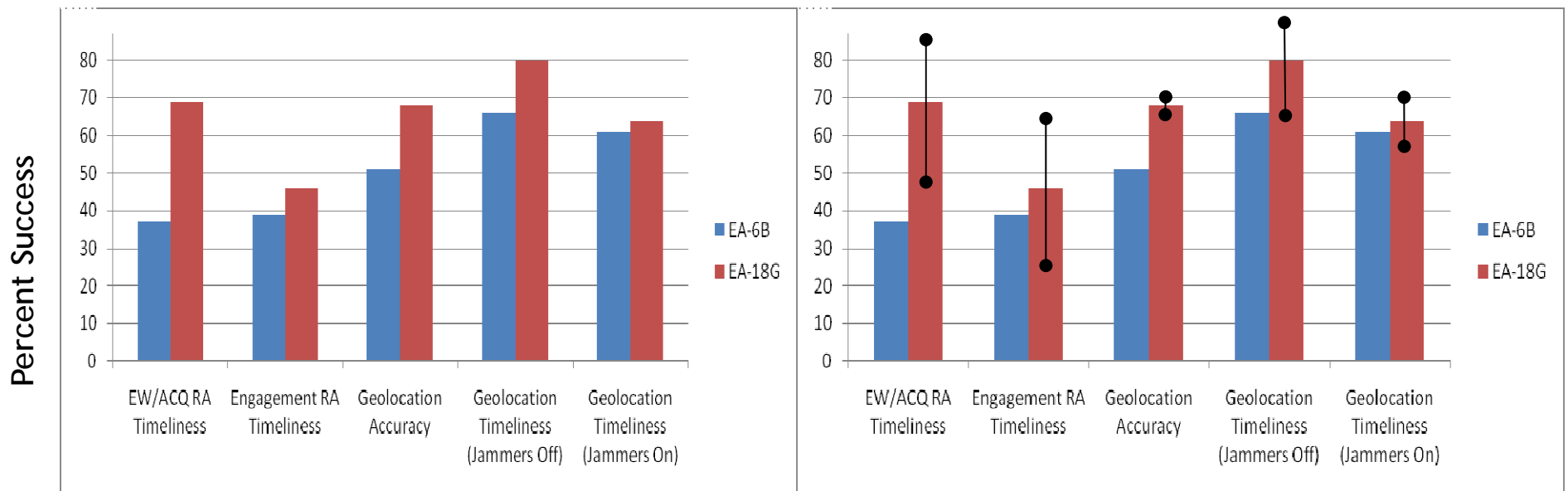
Time to restrict missions




# Improving Operational Testing

# Why did we need to improve test methods?

Figure from DOT&E EA-18G BLRIP



# DOT&E Sets Policy and Guidance for Conducting Operational Testing

 <p>OFFICE OF THE SECRETARY OF DEFENSE 1700 DEFENSE PENTAGON WASHINGTON, DC 20301-1700</p> <p>OCT 19 2010</p> <p>MEMORANDUM FOR COMMANDER, ARMY TEST AND EVALUATION COMMAND COMMANDER, OPERATIONAL TEST AND EVALUATION FORCE COMMANDER, AIR FORCE OPERATIONAL TEST AND EVALUATION CENTER DIRECTOR, MARINE CORPS OPERATIONAL TEST AND EVALUATION ACTIVITY COMMANDER, JOINT INTEROPERABILITY TEST COMMAND DEPUTY UNDER SECRETARY OF THE ARMY, TEST &amp; EVALUATION COMMAND DEPUTY, DEPARTMENT OF THE NAVY TEST &amp; EVALUATION EXECUTIVE DIRECTOR, TEST &amp; EVALUATION, HEADQUARTERS, U.S. AIR FORCE TEST AND EVALUATION EXECUTIVE, DEFENSE INFORMATION SYSTEMS AGENCY DOT&amp;E STAFF</p> <p>SUBJECT: Guidance on the use of Design of Experiments (DOE) in Operational Test and Evaluation</p> <p>This memorandum provides further guidance on my initiative to increase the use of scientific and statistical methods in developing rigorous, defensible test plans and in evaluating their results. As I review Test and Evaluation Master Plans (TEMPs) and Test Plans, I am looking for specific information. In general, I am looking for substance vice a 'cookbook' or template approach - each program is unique and will require thoughtful tradeoffs in how this guidance is applied.</p> <p>A "designed" experiment is a test or test program, planned specifically to determine the effect of a factor or several factors (also called independent variables) on one or more measured responses (also called dependent variables). The purpose is to ensure that the right type of data and enough of it are available to answer the questions of interest. Those questions, and the associated factors and levels, should be determined by subject matter experts -- including both operators and engineers -- at the outset of test planning.</p>	<p>for when I approve TEMPs and</p> <p>evaluation of end-to-end tic environment.</p> <p>es for effectiveness and parameters but most likely there</p> <p>ess and suitability.</p> <p>y, develop a test plan that factors across the applicable levels nation in order to concentrate</p> <p>ss both developmental and interest.</p> <p>ence) on the relevant response tical measures are important to can be evaluated by decision- e off test resources for desired</p> <p>entify the metrics, factors, and nd suitability and that should be</p> <p>reflected in defined test plans. DOT&amp;E is working with other members of the test and evaluation community to develop a two-year roadmap for implementing this scientific and rigorous approach to testing. I am looking for as much substance as possible as early as possible, but each TEMP revision can be tailored as more information becomes available. That content can either be explicitly made part of TEMPs and Test Plans, or referenced in those documents and provided separately to DOT&amp;E for review.</p> <p><i>J. M. Gilmore</i> J. Michael Gilmore Director</p> <p>cc: DDT&amp;E</p>
--	---

☐ The goal of the experiment. This should reflect evaluation of end-to-end mission effectiveness in an operationally realistic environment.

☐ Quantitative mission-oriented response variables for effectiveness and suitability. (These could be Key Performance Parameters but most likely there will be others.)

☐ Factors that affect those measures of effectiveness and suitability. Systematically, in a rigorous and structured way, develop a test plan that provides good breadth of coverage of those factors across the applicable levels of the factors, taking into account known information in order to concentrate on the factors of most interest.

☐ A method for strategically varying factors across both developmental and operational testing with respect to responses of interest.

☐ Statistical measures of merit (power and confidence) on the relevant response variables for which it makes sense. These statistical measures are important to understanding "how much testing is enough?" and can be evaluated by decision makers on a quantitative basis so they can trade off test resources for desired confidence in results.



# Laying the foundations for statistical methods in T&E



Research Consortium

Offsite Meeting

Charter



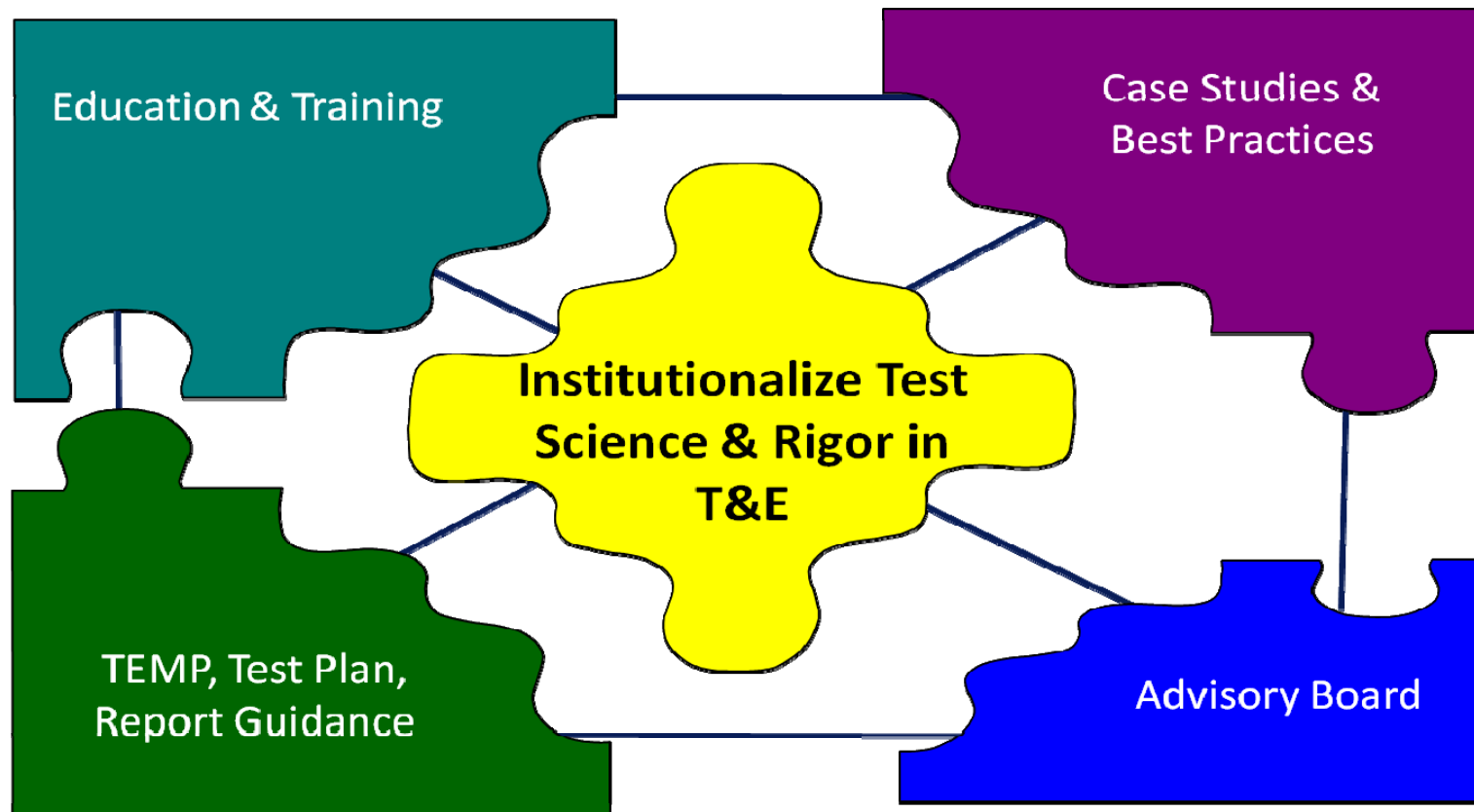
Statistical Engineering with NASA



AO Training, OTA Training





# Puzzled??




# Sharing lessons learned advanced our mutual understanding


UNCLASSIFIED

 OVERVIEW OF  
VALIDATION  
SIMULATION


 **ATEC**

 **A Case Study**  
At

**Case Study**  
**DOE at MCOTEA**

 **MCOTEA**

**F-22 FOT&E 3.1 Test Design**

 **AFOTEC DET 6**

**Stuart Butts**  
**AFOTEC Det 6 Technical Advisor**  
Release Date: 27 Sep 11

DR. STARGEL DOANE  
01B TEST DESIGN AND ANALYSIS  
2/1/2017

**Army Proven  
Battle Ready**

Presented by: Swala Burns  
Written by: Brittney Cates  
Mathematical Statisticians  
May 2011

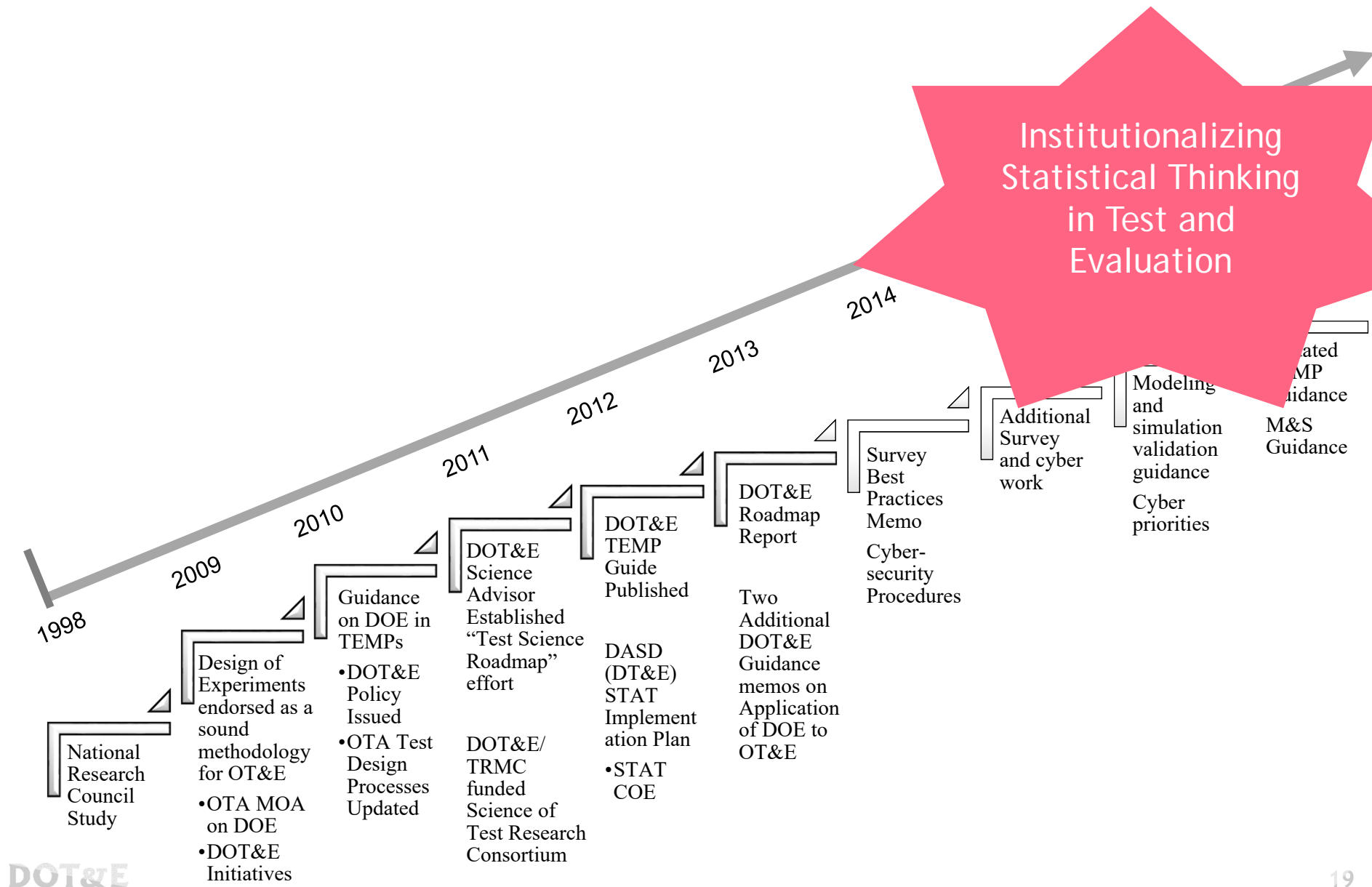
Distribution authorized to Department of Defense and US DoD contractors only; administrative or operational use (Sep 11). Refer other requests for this document to HQ AFOTEC/HQ at 8600 Gibbon Blvd SE, Kirtland AFB NM 87117-4648.

WARNING - This document contains technical data whose export is restricted by the Arms Export Control Act (Title 22, U.S.C., 2751, et seq.) or the Export Administration Act of 1979, as amended (Title 48, U.S.C., App. 2401, et seq.). Violation of these export-control laws is subject to severe criminal penalties. Dissemination of this document is controlled under DoD Directive 5200.26.

This document contains information EXEMPT FROM MANDATORY DISCLOSURE under the FOIA, Exemption 3 applies.

DESTRUCTION NOTICE - Follow the procedures in DoD 5225.22-M, National Industrial Security Program Operating Manual (NISPOM), section 6-706, or DoD 5205.1-R, Information Security Program, chapter VI.

# Without a destination, any path will do



# Lessons Learned from Implementing DOE

Strong leadership

Communicate, communicate, communicate

Find partners

Compromise

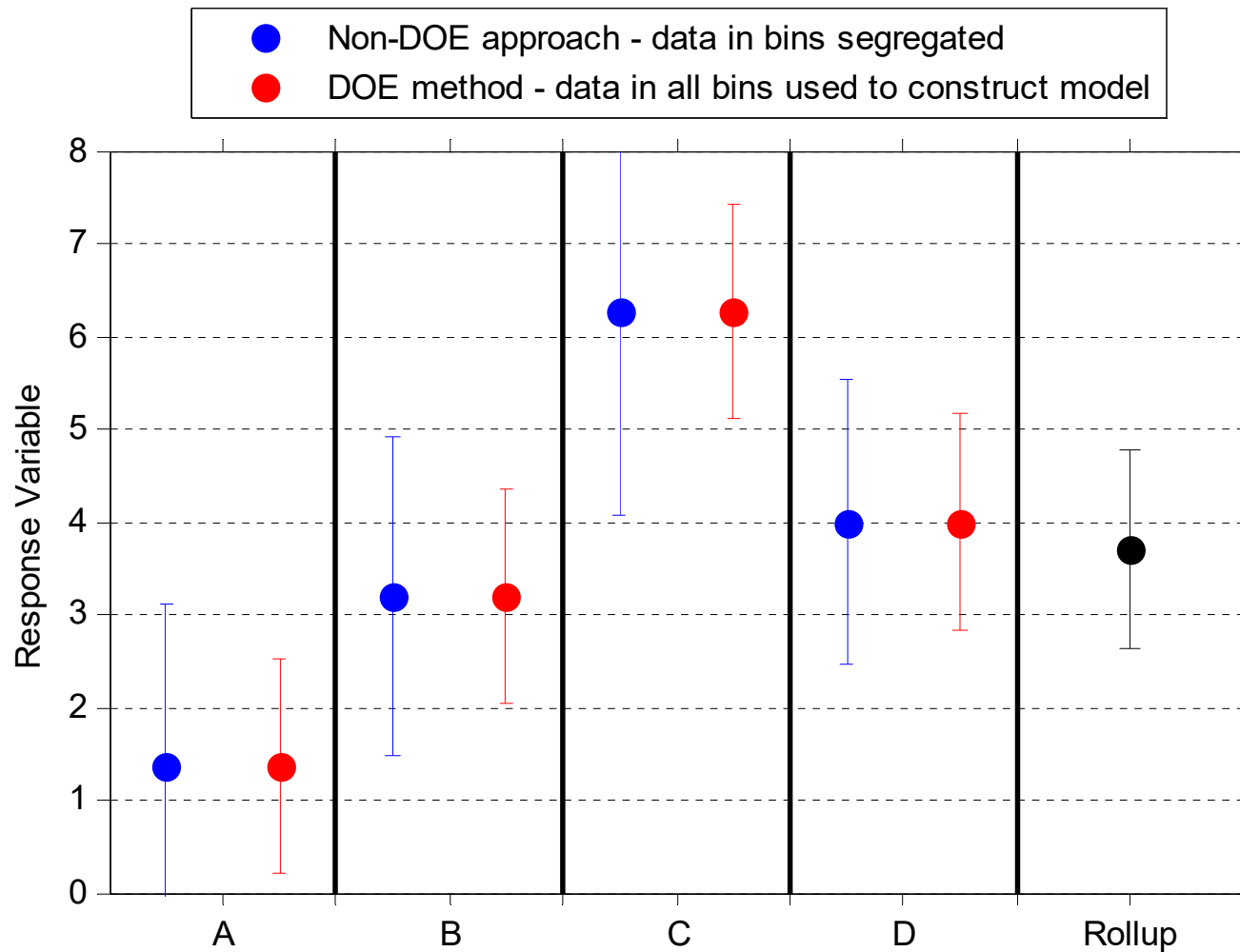
Be open to new ideas

Create quick successes and highlight them

Support the workforce

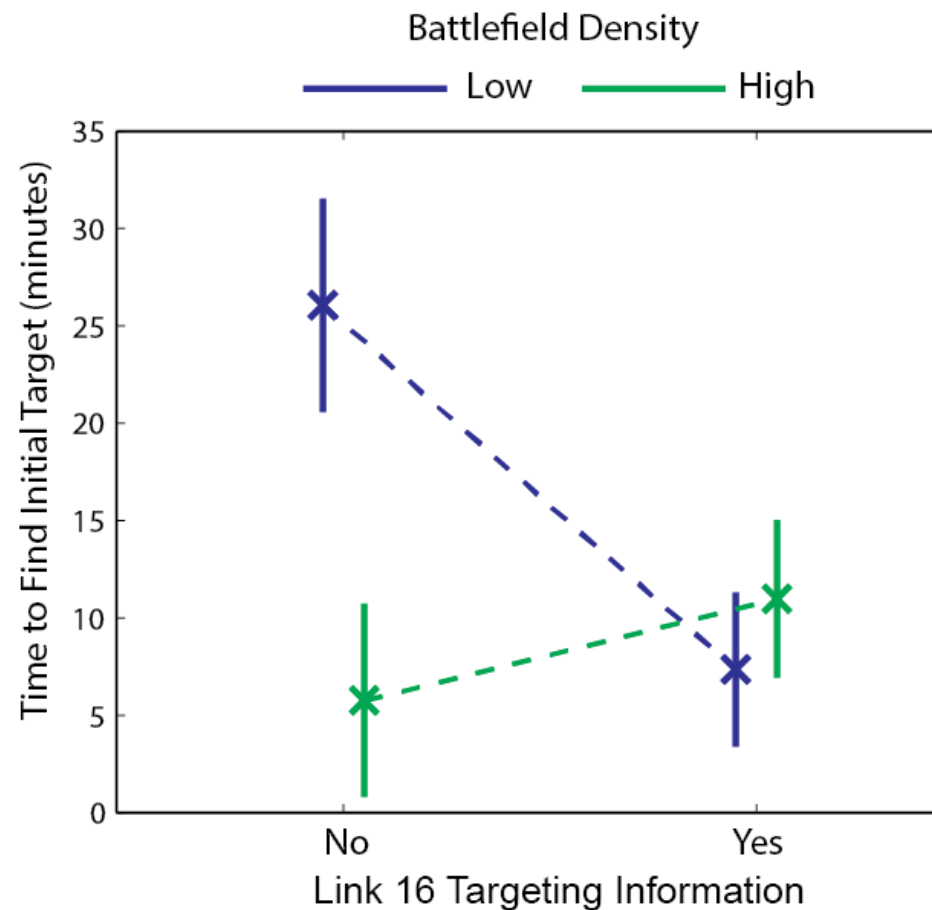
# Statistical Analysis Methods for Improving Mission Characterization

# Statistical analyses maximize information



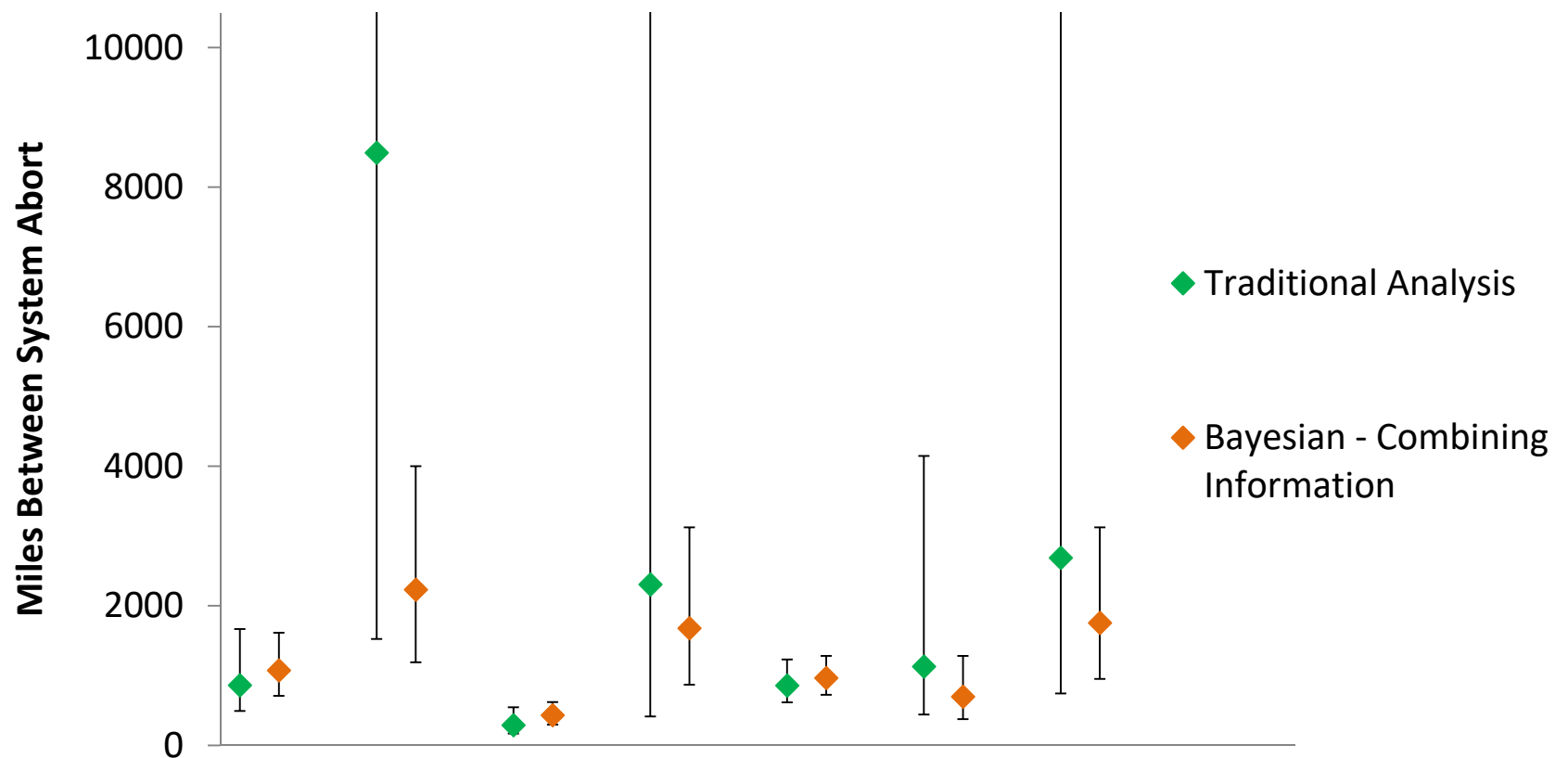


# Statistical models capture important interactions – Apache FOT&E



# Continuing the Path Forward

# Bayesian methods provide flexibility in combining information - Stryker Family of Vehicles Reliability



# Sometimes mission outcome is subjective



Survey regarding improved situational

Strongly agree ☐

Agree ☐

Slightly agree ☐

Slightly disagree

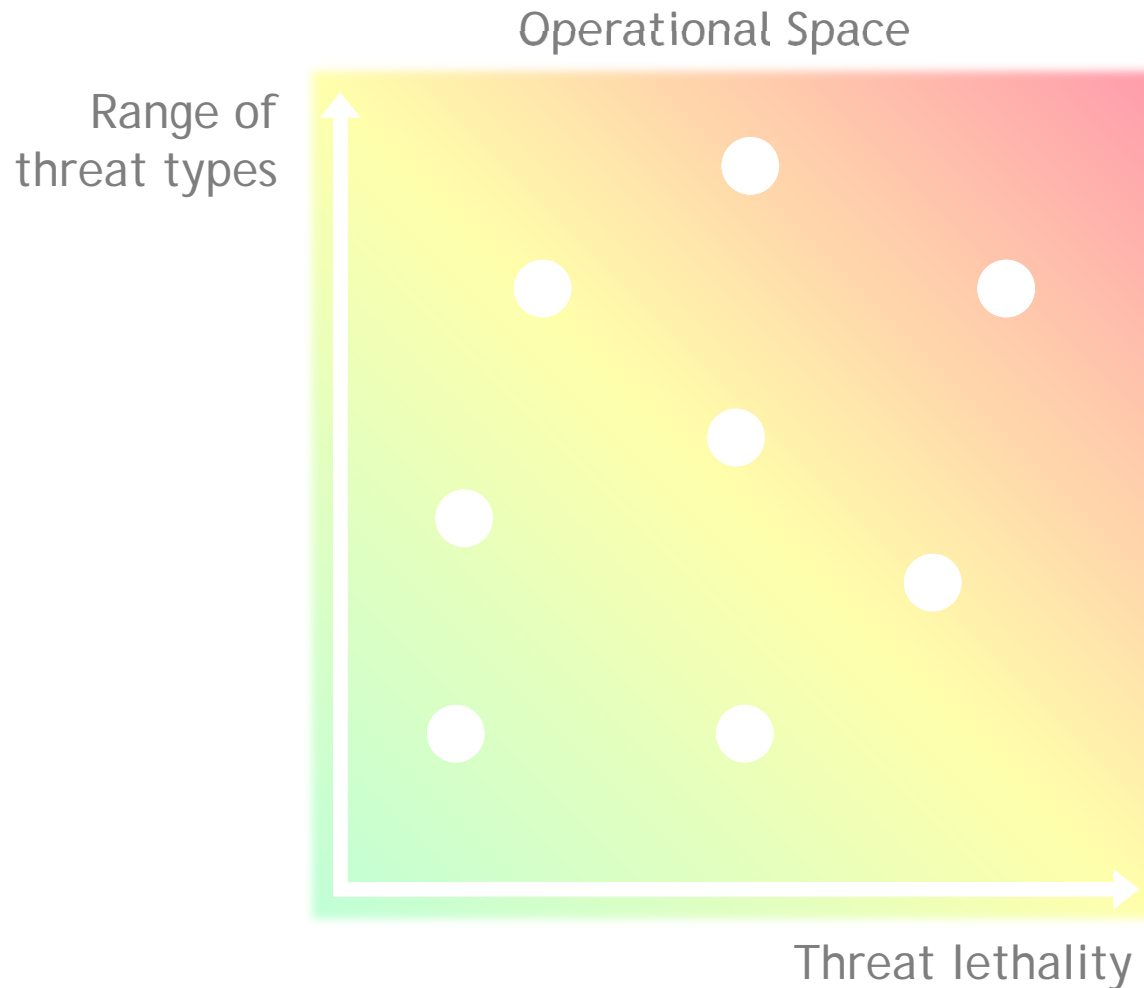
Disagree ☐

Strongly disagree

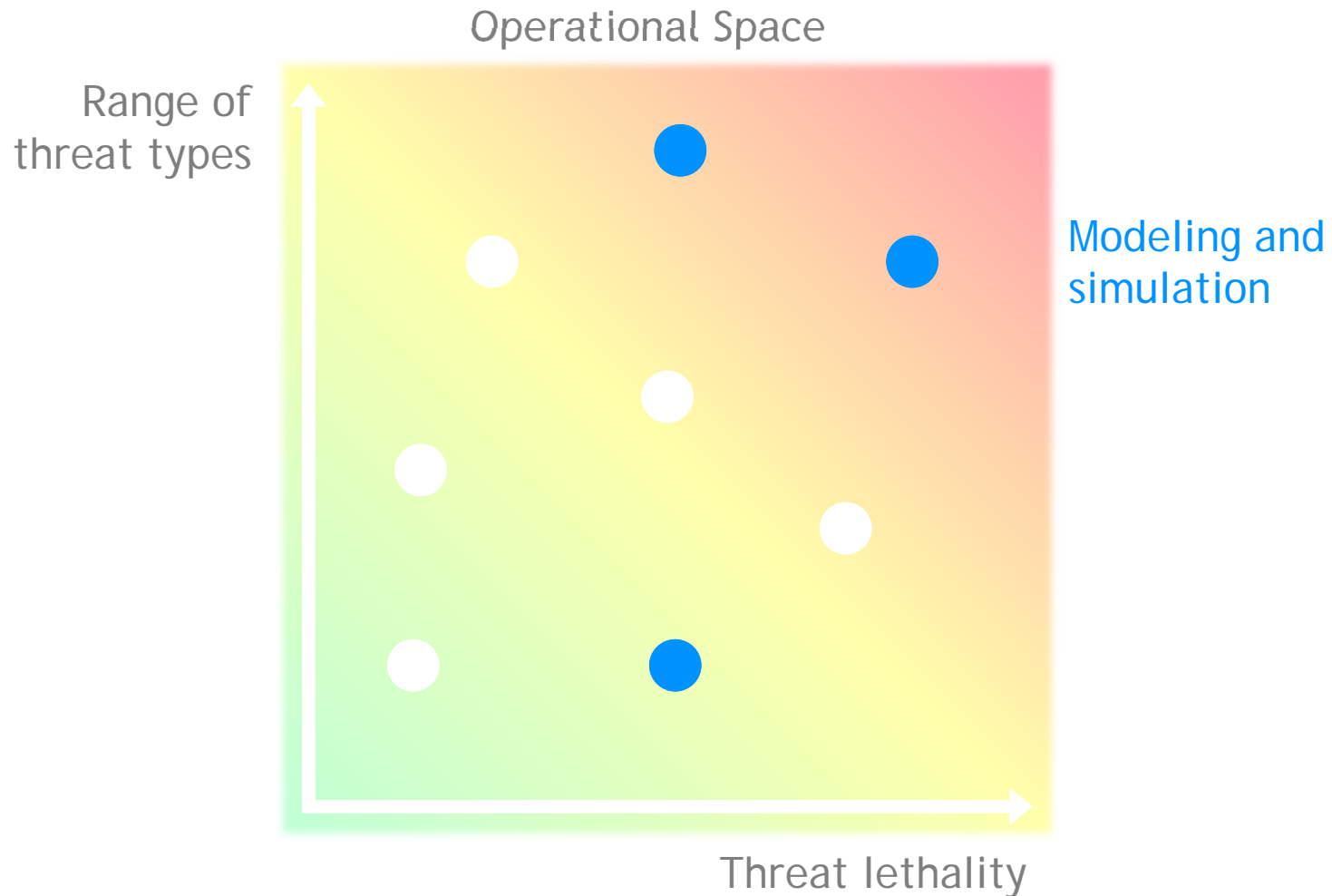
# Guidance highlighted key concepts for improving surveys

- Surveys are appropriate for quantitatively measuring operator and maintainer thoughts and opinions
- Have an administration plan for surveys and only use surveys when appropriate
- Use the right survey
  - Empirically vetted surveys should be used to measure known constructs (e.g., workload, usability, trust)
- Custom surveys should be used appropriately
  - Follow best practices for writing questions
  - Always pre-test
  - Avoid asking questions without a clear analysis plan
- Use interviews and focus groups for problem identification and general context
  - Do not develop lengthy exhaustive surveys about every problem that could occur

# Live-Virtual Constructive simulations can help us learn more... but it needs better validation

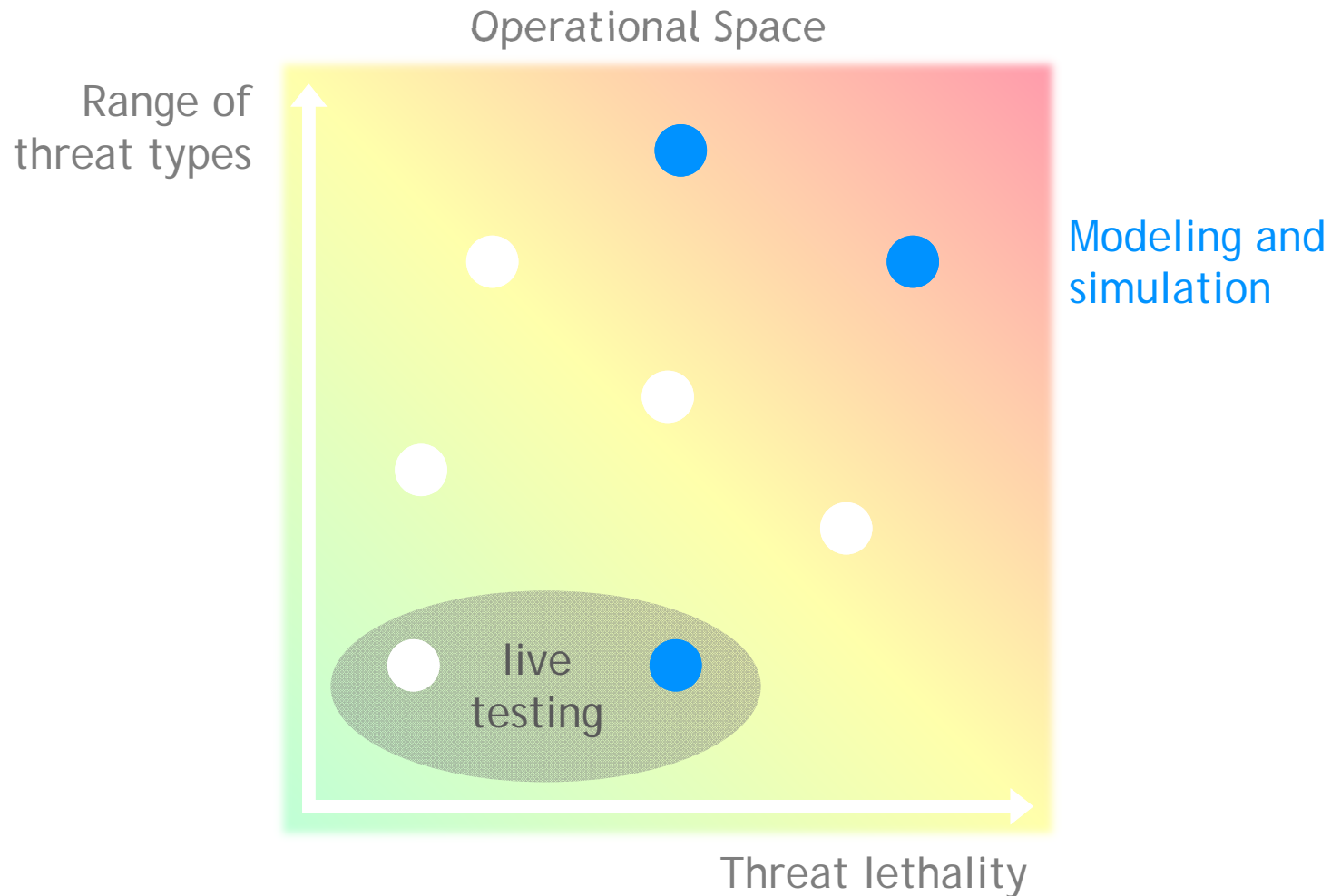


# Live-Virtual Constructive simulations can help us learn more... but it needs better validation

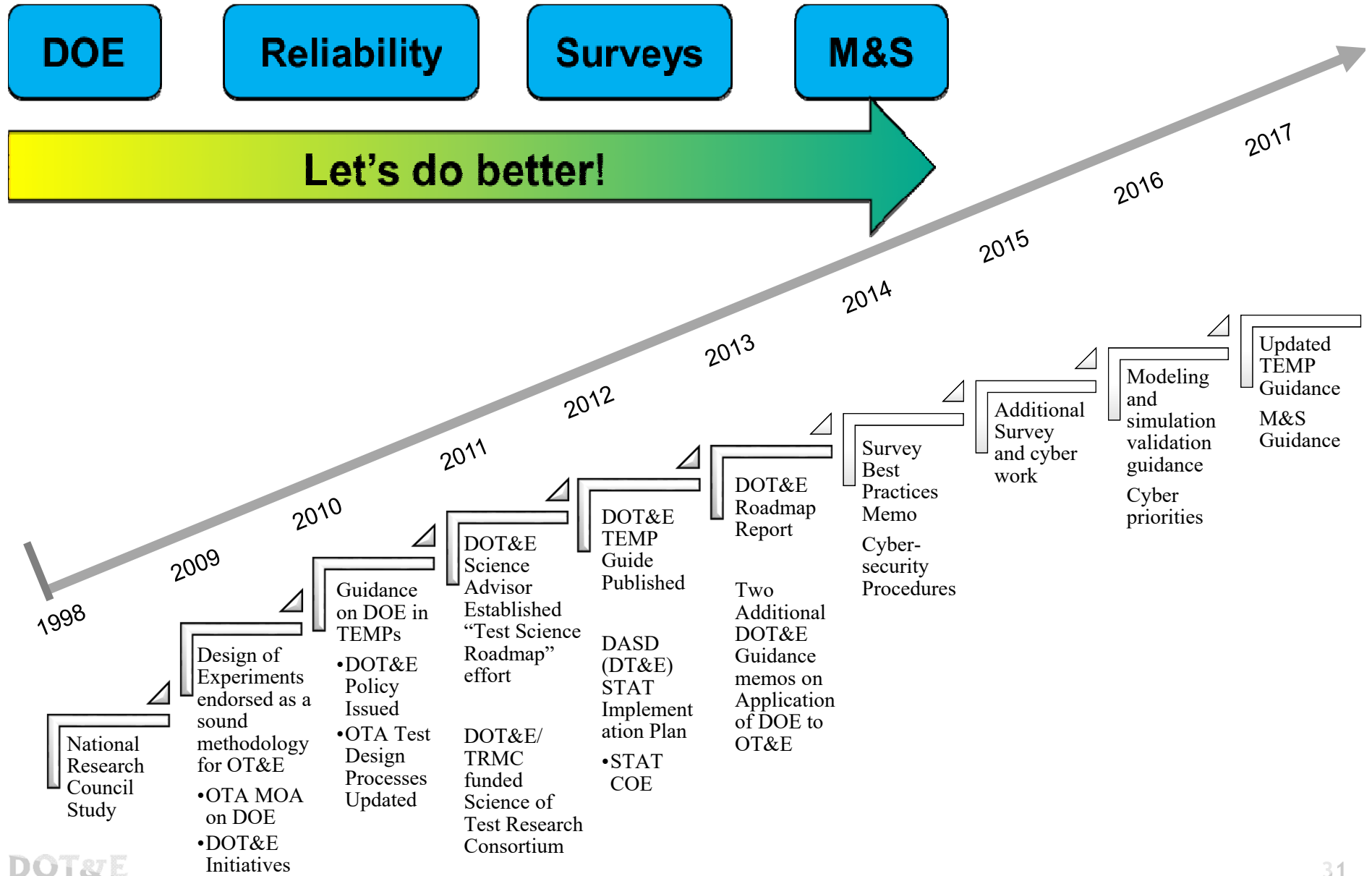




# Live-Virtual Constructive simulations can help us learn more... but it needs better validation

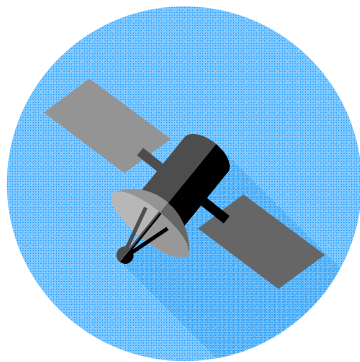


# We continue to increase the statistical defensibility of DoD Test and Evaluation

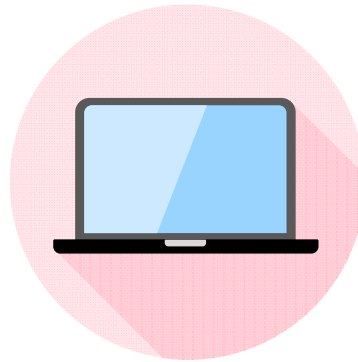


# Future Test Challenges

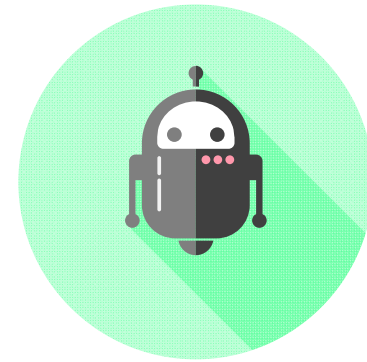
# We need to think carefully about the challenges we face in the future



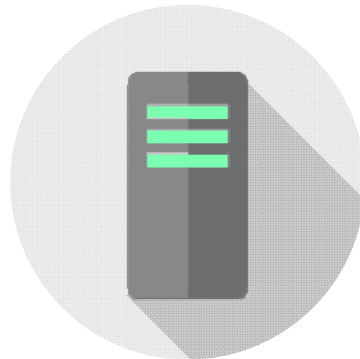
Space



Cyber



Autonomy



Big data



Workforce