

MEASURING THE VALUE OF SOFTWARE ASSURANCE

Prepared for the NASA Aerospace Safety Advisory Panel

Joel Wilf

Technical Staff

Quality Assurance Office

Jet Propulsion Laboratory, California Institute of Technology

October 31, 2013

Topics

- The “Poster” Metric
- The “Bacon Saver” Metric
- Quest for the Value of Assurance
- The SA Value Proposition
- Quantitative Cost and Qualitative Value
- Value Chains to Model SA Benefits
- Quantifying SA Value – An Example
- Metrics for the NASA SAWG
- Where do we go from here?
- Acknowledgements
- Discussion

The “Poster” Metric (1/2)

Flight Software Key Process and Product Metrics			
Process Performance	Effort Growth from PDR	Productivity (Lines of Code/ Work Month)	Defect Density (Defects/ Thousand Lines of Code)
Robust Process	39%	150	4.3
Low to Moderate Process Performance	116%	106	5.9

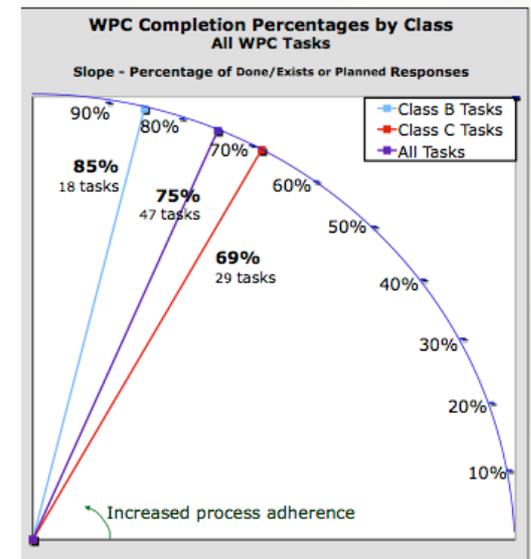
The “Poster” Metric is meant to show projects that the right process rigor will keep costs under control, raise productivity, and lower defects.

Source: Jairus Hihn for the JPL Software Quality Improvement (SQI) initiative

- Works by correlating base measures:
 - Process performance = Percent that project tailoring adheres to standard processes (robust ~ 80%): from tailoring records and work product checklists
 - Effort growth from PDR: from planned vs. actual data
 - Productivity: from SLOC, effort, and schedule data
 - Defect density: from SLOC and defect data in bug tracking systems

The “Poster Metric” (2/2)

- Note: Assurance processes are part of tailoring record and assurance level is highly correlated with “process rigor”
- This metric is a **good indicator**. Why shouldn't SA use it as its value measure?
 - Correlation != Causation
 - Many factors are involved in project benefits: this metric doesn't show **to what extent** assurance (or even “process rigor”) contributes
 - Most important, the poster metric is not **useful** in an operational sense. It doesn't give us information on **how to improve SA** value or tell **how much SA is needed** to achieve the results



Varr-ooooom!

Source: Jairus Hihn,
State of JPL Software
Report

The “Bacon Saver” Metric (1/2)

- The “Bacon Saver” approach wants to equate SA value with the cost the project **would** have incurred if SA hadn’t discovered the defect
- IVV has used this metric
- JPL records “Success Stories” (see opposite) but does not use these as its value metric, because:
 - The measure itself is limited: it doesn’t account for the indirect nature of SA defect discovery, the likelihood of the discovery (by SA or others), or the probability of the defect manifesting as a failure
 - More important: It assumes **SA’s primary function is to find defects (not true)!** This would end up devaluing SA in relation to testing, which *is* the defect-finder

“[Juno] Success Story: SQA review of test completeness led to the discovery of a critical defect”

We shouldn’t say, “This defect could have ended the \$2 billion mission so SA value is \$2 billion!”

The “Bacon Saver” Metric (2/2)

- Treemap, below, shows SA activities (each small box) that result in no findings (green), one or more findings (red), or not covered (gray)
- All too often, we look at just the red boxes and miss the value of assessing products/processes when there are no findings (green)



Source: SA activities and findings in the SQA JIRA issue tracking system

Visualized by Martin Feather

Quest for the Value of Assurance

- In 2009, JPL began a quest to determine the value of SA
- Recruited help from Professor Dan Port, University of Hawaii
- Began with detailed survey of SA stakeholders on their perception of SA value and their “win conditions”
- Initial results:
 - Stakeholders were often confused about the definition of SA
 - Stakeholders had different win conditions (see list, opposite)
- Realization: We needed a “value proposition” for SA that...
 - Didn't try to satisfy all win conditions
 - Gave unified way to assess SA value
 - Connected SA as “a set of activities” to SA as an “umbrella risk reduction strategy”

Win Condition

Ensure complete compliance

Discover quality defects in work products

Certify SW for I&T/ATLO/Flight

Uncover hidden SW risks

Ensure testing correct and complete

Fix SW quality problems

Early defect detection

“Win Conditions” identified in survey of various SA Stakeholders

Source: Port, D.; Wilf, J., "A Study on the Perceived Value of Software Quality Assurance at JPL," *System Sciences (HICSS)*, 2011 44th Hawaii International Conference on , vol., no., pp.1,10, 4-7 Jan. 2011

The SA Value Proposition

*Software Assurance enables more **confident** decision-making by providing independent credentialed **information** to reduce uncertainty in systems decisions that depend on quality – and thus **reduces decision risk***

- Implications to SA practice:
 - SA engineers are not the “process police”
 - SA provides information, findings, and associated risk...
 - But projects make the decisions and mitigate/accept risk
- Connecting SA value to decision risk brings SA into the realm of economic decision theory
 - Puts a dollar figure on value
 - Enables consideration of cost vs. value
 - Enables answering “how much assurance is enough?”
 - Guides us in evaluating improvements in SA practice

Quantitative Cost + Qualitative Value (1/3)

- Using value proposition with Cost Model to budget:
 - SA Activity-based Cost Model
 - Per-Activity Basis of Estimate (BOE) data (e.g., requirements/hour, pages/hour, etc.)
 - Initial data based on SA memory (updated using actuals)
 - Cost model guides SA discussion with project
 - Discussion focuses on cost vs. project need/value of SA
 - Quantitative cost + qualitative value
- Using value proposition to respond to descopes
 - SA Activity-based Descope Table
 - Per-Activity examination of cost vs. impact of loss of activity
 - Descope table guides SA discussion with project
 - Quantitative cost + qualitative value discussion
- Future: integrate descope table into Cost Model
- Future: quantify value information

Quantitative Cost + Qualitative Value (2/3)

	# of Meetings/ Reviews	Hours Per Action	Total HOURS
6. Project Review Meetings/Major Milestone Reviews Participation:			
A - Participate: Monthly Management Reviews (MMRs) @ 4 hours each (prepare, attend, follow-up) (FPP 5.16.4) - Assume 4 hours/month (OSMS support covered above); Project/LM Spacecraft MMR	30	4	120
B1 - Participate: Participate and support project team meetings Assume 134 weekly meetings at 2 hour per meeting	134	2	268
B2 - Participate: General SQA program support - unscheduled meetings and requests for assistance not covered by other categories. SQA assumes 10 per cent of time = 4 hours per week.	132	1	132
C1 - Participate: Participate in major project milestones (PMSR, PDR, CDR, ARR, etc.) (SDR 4.2.4): 40 hours per milestone review: Assume CDR, SIR, Pre-Ship Review, COFR; S/C FSW PDR, CDR	4	40	160
C2 - Participate: Participate in major instrument milestones (PDR, CDR.) (SDR 4.2.4): 40 hours per milestone review: IDA and SEIS instrument FSW; Spacecraft FSW	6	40	240
C3 - Participate: Participate in major GDS milestones (TRR, DDR): Assume 5 deliveries/40 hrs per delivery	5	40	200
D - Participate: Project Inheritance Reviews (FPP 6.6.3)(SDR 4.2.9): 40 hours per system: Robotic Arm	1	40	40
E - Participate: Project Engineering Reviews (e.g., Peer Reviews) (FPP 6.6.3)(SDR 4.2.9): 8 hours/review - Assume 20 reviews total	20	8	160
Total Review Meetings/Major Milestone Review Support:			1320

Cost Model: Cost/BOE metrics + value discussion with project

Quantitative Cost + Qualitative Value (3/3)

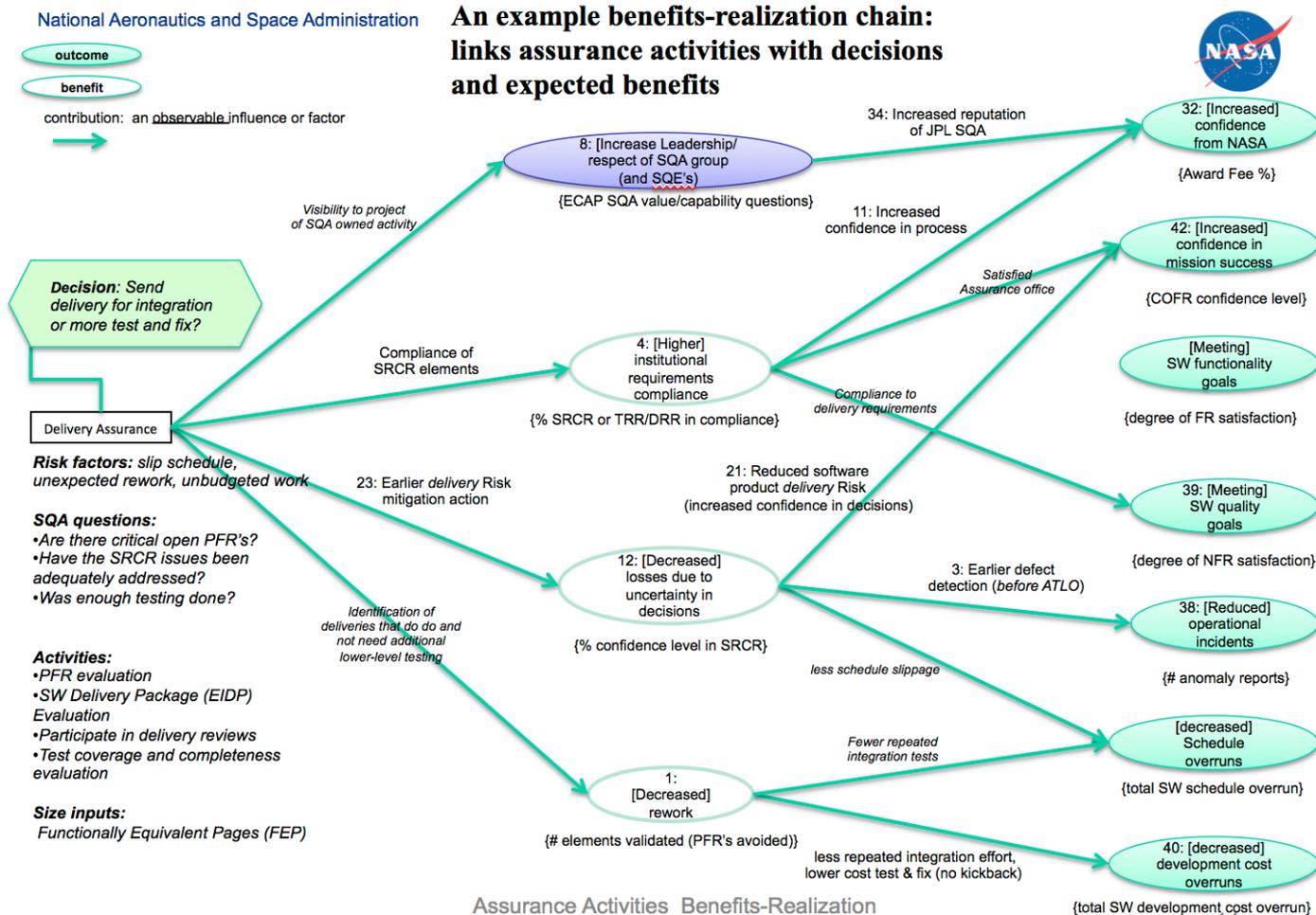
Activity	Effort (hours)	Results	Decisions Supported	Expected Benefit	Potential Impact if Descoped
Audit Subcontractor Artifacts	Not budgeted	Identify non-compliances, rate associated risks, & identify/accept mitigations	Suitability of software processes for given tasks:	Identification of inadequate processes leading to defective products.	1. Creation of CMMI and ISO findings & potential loss of CMMI L3. 2. Increased chance of defective products.
Review FSW & GDS Code (reduced level of effort)	1864	1. Identify defects 2. Identify improvement opportunities	Suitability of software quality	1. Increased confidence that software will work as intended. 2. Increased confidence that software will recover from errors.	1. Mission critical defects will remain, potentially leading to mission failure. 2. Unnecessary performance/resource limitations may exist.
Review ICDs & FDDs	840	1. Identify implicit requirements that should be flowed down to software engineering (e.g. interface protocols) 2. Identify software hazards	1. Acceptance of requirements 2. Acceptance of architectural design with respect to safety	1. Increased confidence that requirements are properly allocated to software. 2. Identification of software hazards in time to mitigate them through changes in requirements and architecture.	1. Will not be able to perform software hazard analysis. 2. Lower confidence that requirements are complete. 3. Greater likelihood that hazards will remain in the software.
Audit MOS, SDS, SSE Artifacts	272	Identify non-compliances, rate associated risks, & identify/accept mitigations	Suitability of software processes for given tasks:	Identification of inadequate processes leading to defective products.	1. Creation of CMMI and ISO findings & potential loss of CMMI L3. 2. Increased chance of defective products.
Conduct Delivery Reviews (reduced level of effort)	192	Identification of unmet or untested requirements or inadequate delivery documentation.	Suitability of delivery to System I&T	Improved readiness of delivery	Increase in rework

Descoped tables: Relating budget loss to decisions and risk

Value Chains to Model SA Benefits (1/3)

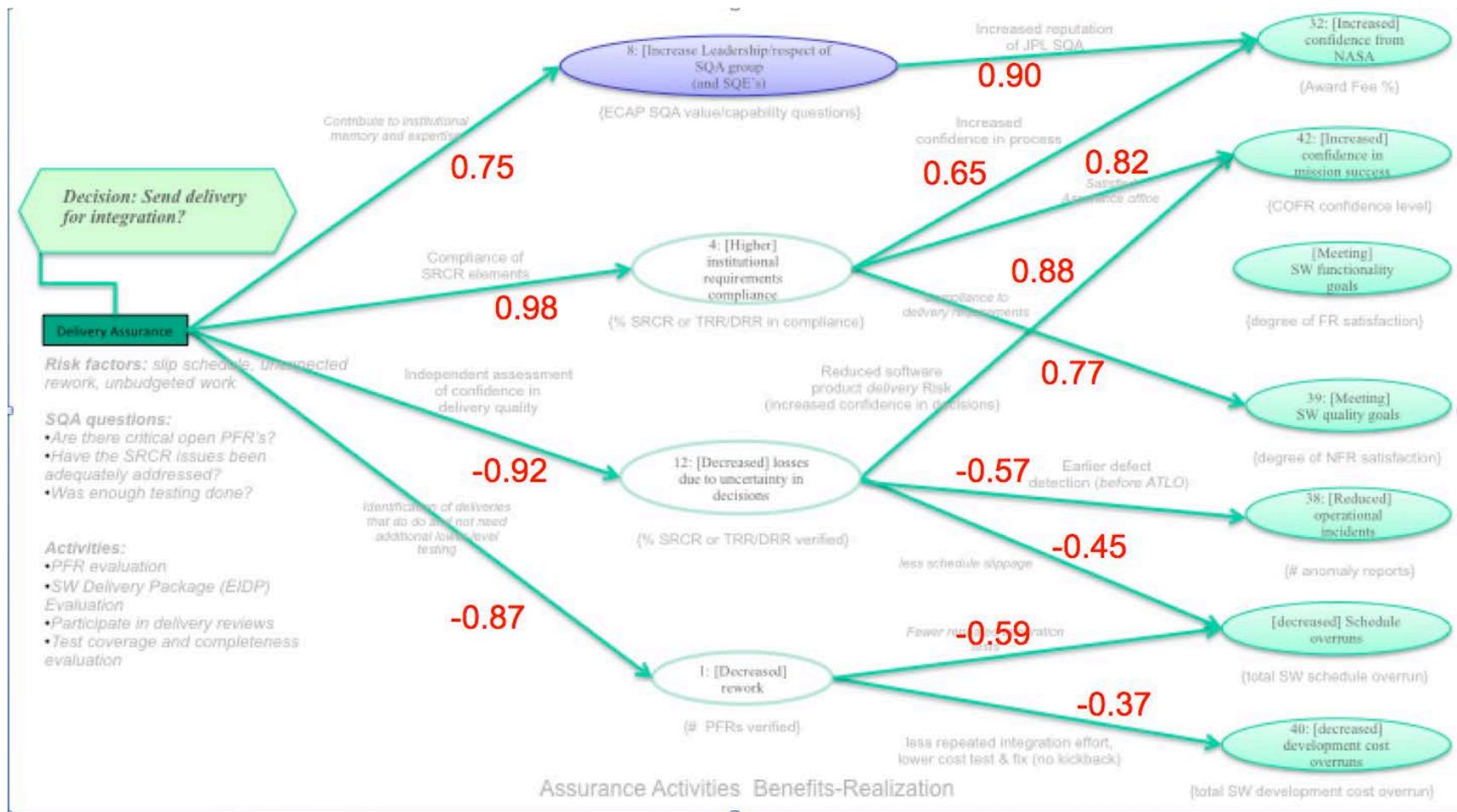
- In addition to decision risk reduction, there are other benefits of SA (e.g., lower cost, fewer defects, etc.)
 - Expanding on idea of “poster” metric discussed above
 - But taking into account that SA may be one of many factors leading to the desired benefit
- Value (or Benefit Realization) Chains
 - Model outcomes and resulting benefits for each SA activity
 - Graph a causal chain, starting from activity
 - May be used qualitatively
 - Techniques for estimating the “strength” of the connection
- Dan Port created Value Chain for SW delivery assurance
- Future: develop value chains for all main SA activities

Value Chains to Model SA Benefits (2/3)



Value Chain for Software Delivery Assurance (SRCR)

Value Chains to Model SA Benefits (3/3)



Value chains with numbers representing degree of influence

Quantifying SA Value – An Example (1/5)

- Goal: Quantify the value of certifying a Mars Rover (MSL/Curiosity) software release using JPL's Software Release Certification Record (SRCR) process.
 - **Warning: This gets a little technical.** The high-level take-away is that quantifying value is possible. Even more detail is available in the following paper:
 - Port, D.; Wilf, J., "The Value of Certifying Software Release Readiness," *Empirical Software Engineering and Measurement*, October 10-11, 2013, Baltimore Maryland
- Step 1: Start with the decision: Send / Don't Send MSL Rover Flight Software (RFSW) to Systems I&T – and potential losses from the decision:
 - Case 0: Send and it passes (loss = 0)
 - Case A: Send and it fails (loss = A)
 - Case B: Hold it and it passes (loss = B)
 - Case C: Hold it and it fails (loss = C)
 - Assume for discussion $A < C < B < 0$

Loss Table	Send	Hold
Pass	0	B
Fail	A	C

Quantifying SA Value – An Example (2/5)

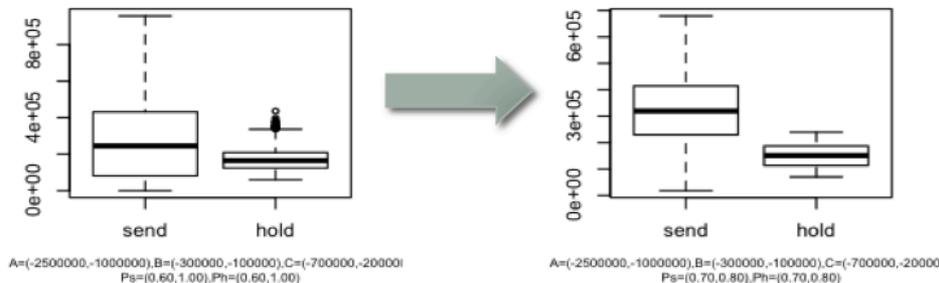
- Step 2: Translate the table of outcomes to one of Opportunity Loss (OL), the loss from making the wrong decision, compared to making the right one.
 - Our SA value assumption: SA reduces the risk of making the wrong decision
 - From the OL perspective, if we send it and it passes, we made the right decision and if we hold it and it would have failed, we made the right decision. The other decisions are wrong and have the following OL:

OL	Send	Hold
Pass	0	-B
Fail	C - A	0

- If p is the probability of passing I&T, then the OL table gives us a strategy for deciding whether to Send or Hold, based on minimizing Expected Opportunity Loss (EOL):
- If $p*(-B) > (1-p)*(C - A)$ then the EOL of Hold $>$ EOL of Send, and we should Send

Quantifying SA Value – An Example (3/5)

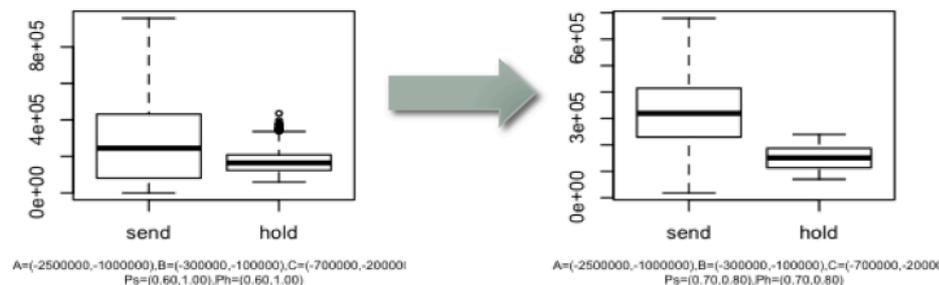
- If the project knew the values of p , A , B , and C exactly, they could just apply the formula every time – and not need SA!
- But p , A , B , and C are always uncertain; the project is always making decisions under uncertainty
- Step 3 is estimating with uncertainty the values of p , A , B , and C prior to the SRCR review, then the degree to which answering the questions on the SRCR form increases certainty (decreases the range of possible values). For example, if the SRCR reduces the uncertainty of passing (p) from the range (.6, 1) to (.7,.8), the EOL for Send vs. Hold changes as follows:



EOL for Send vs. Hold decisions, before and after the delivery assurance has reduced uncertainty

Quantifying SA Value – An Example (4/5)

- Note: Estimate of uncertainty with “credible intervals”
 - These are basically confidence intervals derived from a combination of expert judgment and empirical data
 - For example: Start with most confident range for value to be estimated, i.e. 100% certain, $p = (0,1)$. Now trade confidence for more accurate range based on evidence. For example, if we have data that shows “at least 7 of the 10 releases we thought would pass had passed, and we were willing to be wrong 10% of the time, we could say with 90% certain that $p=(.7,1)$ ”
- Step 4 Using Monte Carlo simulation, create a distribution of Decision Risk, given the uncertainties above
- Less overlap in the Send vs. Hold EOL => less decision risk:



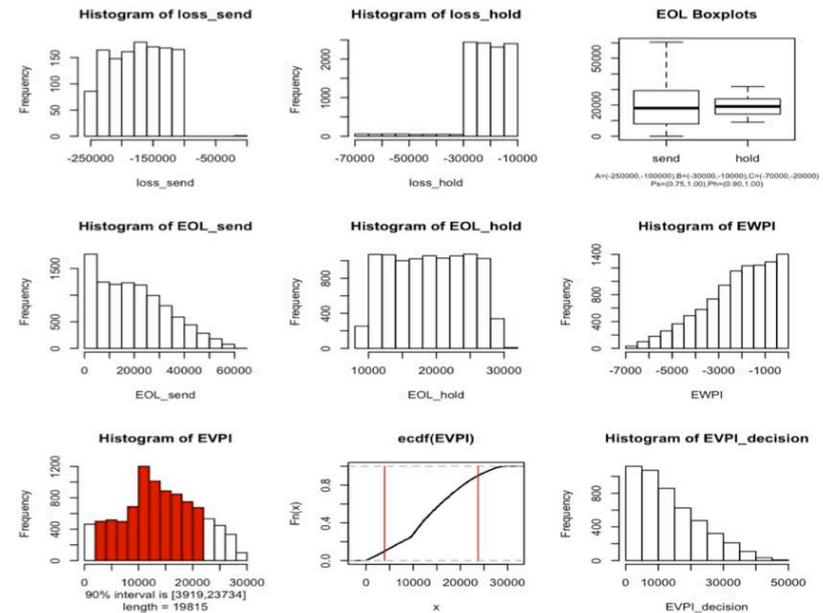
Lower EOL for “hold” with less overlap after delivery assurance means we can make the Hold decision with less decision risk

Quantifying SA Value – An Example (5/5)

- Finally, it's possible to extract a dollar amount from the decision risk distribution: "95% Value at Risk (VaR)" means there is a 95% chance that we will lose no more than VaR amount (due to decision loss)
- That's the "value of assurance"

Metrics for the NASA SAWG (1/2)

- NASA SAWG established a set of strategic goals
- Strategic Goal #2: “Establish a core set of SA performance measures for all Centers across the Agency”
- The Goal #2 Team suggested metrics based on experience with:
 - Various value metrics in use or proposed
 - Base measures that are or should be available at each NASA center
 - Deeper exploration of an SA Value Proposition



Analyses of EOL from decisions

Source: Dan Port, Joel Wilf, Paper on Value of Certification

Metrics for the NASA SAWG (2/2)

Questions	Measures that Answer each Question	How hard?
How well does SA cover NASA projects as required?	Number of projects covered / total potential projects (by attributes such as mission class, software class, safety criticality, software size, etc.)	Easy
How much effort does SA contribute?	SA budget / total project budget (by project attributes)	Easy
How do stakeholders view SA contribution?	Likert Scale Stakeholder survey of attitudes about SA (by stakeholder type, project type, extent of SA effort, SA activity)	Mid
Does SA contribute to reduced defects/failure?	Defects by phase vs. SA effort, failures in operations vs. SA effort	Mid-Hard
What is SA's contribution to all desired project outcomes?	SA Value Chains (SA activities, outcomes, benefits)	Hard
What is SA's contribution to key project decisions?	Decision opportunity loss (per decision, see Port/Wilf paper)	Hard

Where do we go from here?

- Where are we now?
 - Some understanding of existing SA metrics
 - Good baseline measures to work with: SW process tailoring, SW cost/actuals, SA cost/actuals, defects of various types, findings, SA user surveys
 - Exploration of the meaning of the value of SA and how we might measure it
- What could we do?
 - Integrate value/risk information into SA cost models
 - Create ability to understand “how much SA is enough”
 - Create ability to manage and improve SA with metrics
- What do we need to get there?
 - Executive commitment
 - Resources

Acknowledgements

- Professor Daniel Port, University of Hawaii – for insights connecting SA value to economic decision theory, the SA value proposition, and continuing work in measuring the value of SA
- NASA SAWG and SARP – for encouraging and supporting SA metrics
 - Martha Wetherholt – NASA SA Technical Fellow and Metrics Goal sponsor
 - Wes Deadrick – SARP Manager and potential metrics sponsor
 - Lisa Montgomery – Former SARP Manager sponsored metrics-related tasks
- JPL Software Quality Improvement (SQI) initiative – for defining, collecting, analyzing, and archiving JPL institutional metrics
 - Scott Morgan – SQI manager and metrics sponsor
 - Jairus Hihn – Cost, risk, and defect metrics expert
- JPL Software Assurance Researchers
 - Martin Feather – Data visualization expert
 - Allen Nikora – Anomaly data analysis and data mining expert
- JPL Software Assurance (“SQA”) Team – for adopting the value-based assurance mind-set and collecting cost/value data in SA cost models, actuals, activities, findings, success stories, and delivery certifications (SRCRs)
 - Tuan Do – Group Supervisor

Discussion

