Let's Go Agile! Data-Driven Agile Software Cost and Schedule Models

Dr Wilson Rosa
Sara Jardine
Chad Lucas

2022 ICEAA Workshop

May 2022

Presented at the 2022 ICEAA Professional Development & Training Workshop - www.iceaaonline.com/pit2022

Agenda

Introduction

- Problem Statement
- Proposed Solution
- Breakthroughs
- History Behind Agile at DHS

Agile Project Dataset

- Data Collection
- Data Sources
- Variables
- Data Normalization
- Demographics
- Descriptive Statistics

Effort and Schedule Models

- Benchmarks
- Effort Estimation Models
- Schedule Estimation Models
- Model Rankings

Results

- Model Usefulness
- Model Limitations
- Main Takeaways

What is the Problem?

"The Department [DHS] needs a credible and accurate method for estimating the cost of software development programs that can be tracked over time and provide insight into whether a program is behind schedule or is forecasted to exceed initial cost projections."

Stacy Marcott
Acting Chief Financial Officer
May 30, 2019

What is the Solution?

- Offer a set of **data-driven** software development effort and schedule estimating models for DHS agile projects

- Acquisition cost community can use these models to:
  - ✓ Estimate effort and schedule to support DHS and DoD decision reviews of agile programs
  - ✓ Crosscheck vendor proposals and evaluate contractor performance

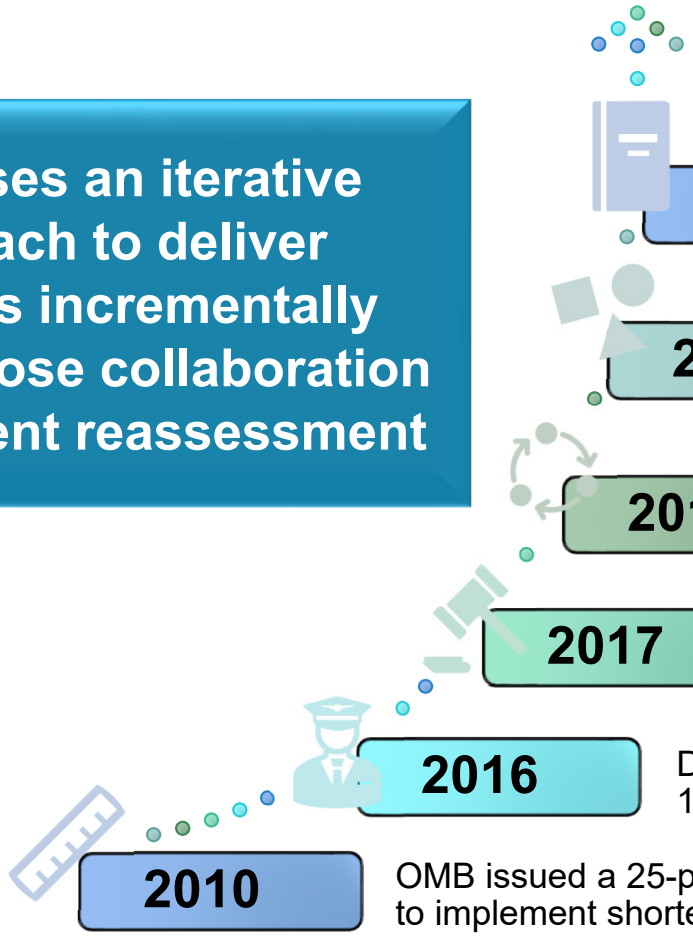
Study Breakthroughs

- Delivers first-ever agile software cost dataset (n=18) for DHS cost community
- Presents a new process for collecting, normalizing, and analyzing agile project cost and schedule data for Firm Fixed Price and Time & Materials contracts
- Introduces ***Functional Story*** as a new sizing measure for agile cost estimation
- Offers **data-driven** agile software project **effort and schedule benchmarks** and **regression models** for **six** different sizing measures:

1. Functional Story
2. Unadjusted Function Point
3. Simple Function Point
4. Story
5. Story Point
6. Issues

History Behind Agile at DHS

Agile **uses an iterative approach to deliver solutions incrementally through close collaboration and frequent reassessment**



2021 DHS Agile Guidebook

2019 DHS CAD launched the cross-agency ***Joint Agile Software Innovation*** (JASI) Cost IPT

2018 DHS Agile Methodology for Software Development and Delivery for IT, Policy Instruction 102-01-004

2017 DHS USM directed DHS CAD to find ways to improve agile software development programs [1]

2016 DHS Agile Development and Delivery for IT Instruction Manual 102-01-004-01

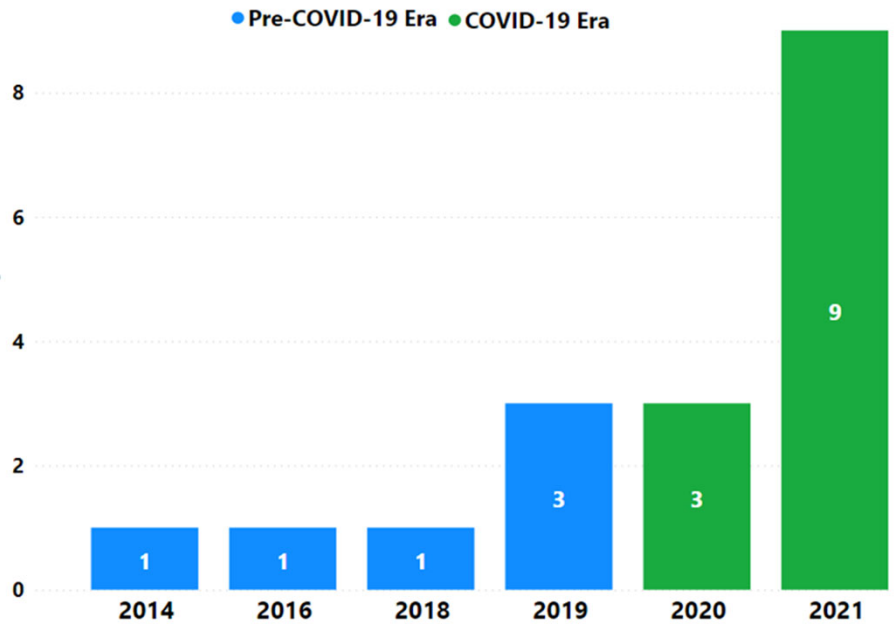
2010 OMB issued a 25-point plan to reform IT projects and called on federal agencies to implement shorter delivery timeframe

Agile Project Dataset

Presented at the 2022 ICEAA Professional Development & Training Workshop: www.iceaaonline.com/pit2022

Data Collection

- 100% data collection efforts occurred during the COVID-19 pandemic
  - March 2020 to January 2022
- Data provided by the Program Managers
- Dataset included **18 agile projects**
  - DHS (15) and DoD (3)
  - Across 11 different companies
  - 12 completed during COVID-19 era

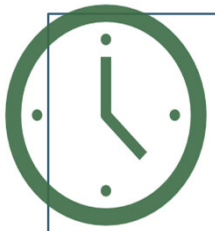


Data Sources

- All data in this study were provided by the Agile Program Management Offices
- 100% obtained from Official/Authoritative Documents:

Effort

- Monthly Contractor Invoices
- Product Backlog



Schedule

- Monthly Contractor Invoices
- Product Backlog (in JIRA)

Size

- Requirements Traceability Matrix
- Functional Requirements Document
- Product Backlog (in JIRA)

Context

- Acquisition Documents
- Agile Core Metrics

Variable Selection (Common Sense)

Dependent

Independent

Categorical

Effort

- Actual labor hours to complete all contractor development activities
- Reported at the release level

Schedule

- Actual development time (months) to complete all software development activities
- Reported at the release level

Functional Story

- Subset of functional requirements describing what the software does in terms of tasks and services

Issue

- Unit of work traced through a workflow, from creation to completion
- Total issues are the sum of stories, bugs, tasks, epics, and others

Story

- Feature or unit of business value that can be estimated and tested
- Describes work that must be done to deliver a feature for a product

Story Point

- Unit of measure to express the overall size of a story, task, or other piece of work in the backlog

Unadj. Function Point

- Function point count without the assignment of complexity to any of the objects counted

Simple Function Point

- Method for sizing software requiring the identification of elementary processes and logic files to approximate a function point count

Scope

- Categorical variable indicating whether the scope of project is an enhancement or full development

Data Normalization:
How did we measure effort?

- Effort hours in this study captures total labor incurred by the contractor's agile development teams
- Total labor includes 11 cost elements aligned to the DHS IT Work Breakdown Structure (WBS)

ID	DHS IT WBS Element
1.i.1	Program Management
1.i.2	Systems Engineering
1.i.4.2	**Software Development**
1.i.4.3	Data Development & Transition
1.i.4.5	Training Development
1.i.4.6.1	Development Test & Evaluation
1.i.4.6.1	Cybersecurity Test & Evaluation
1.i.4.7	Logistics Support Development
1.i.7	System Level Integration & Test
1.i.8.6.1	Help Desk/Service Desk (Tier 3)
1.i.8.6.4	Software Maintenance

Why use total labor?

Reporting labor at the total level (as opposed to software development alone), is recommended since most DHS agile development contracts are FFP or T&M, and generally do not breakout effort by major cost elements as seen in traditional cost-plus contracts

Data Normalization:
Counting Functional Story, SiFP, UFP

Extract Product Backlog (JIRA)

- Go to column titled, *Issue Status* and filter by rows marked as *Done*

Issue Status
Deferred
Done
In-Progress

Step 1

Find Story

- Go to column titled, *Issue Type* and the rows marked as *story* in this column

Issue Type
Other
Task
Story
Bug

Step 2

Find and Count Functional Story

- Categorize each story as *functional* or *non-functional**
- Count rows marked as *functional* **

Category
Functional Story
Non-Functional Story

Step 3

Calculate Function Point

- Convert each functional story into SiFP* & UFP*

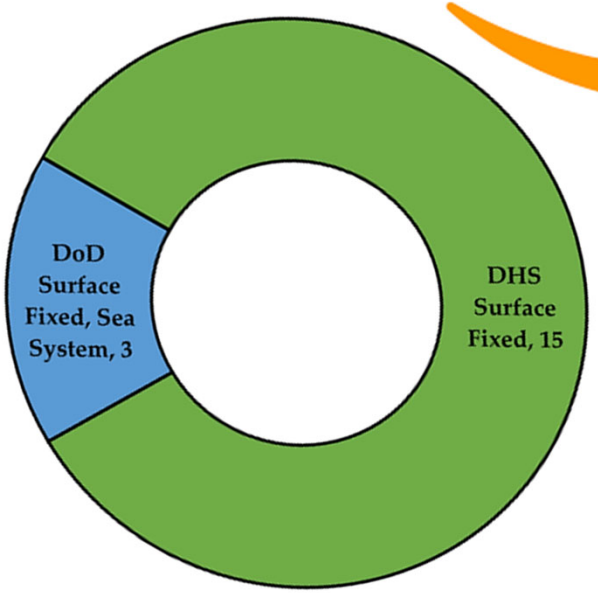
Calculation
SiFP
UFP

Step 4

Notes:
*Performed by a Certified Function Point Specialist
**Functional Stories (from product backlog) = Functional Requirements (from RTM or FRD)

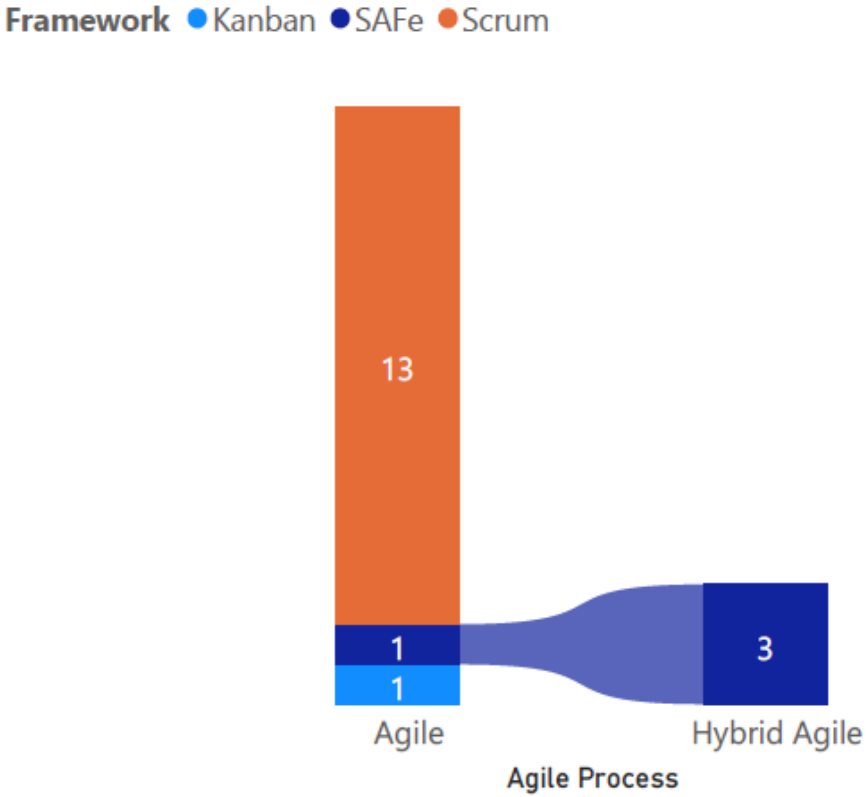
Dataset Demographics

- Sample Size:18 Projects
- Automated Information System
- Majority (13) used cloud-hosted Amazon Web Services
- Majority (15) used FFP or T&M Contracts
- 2 to 4-week Iterations

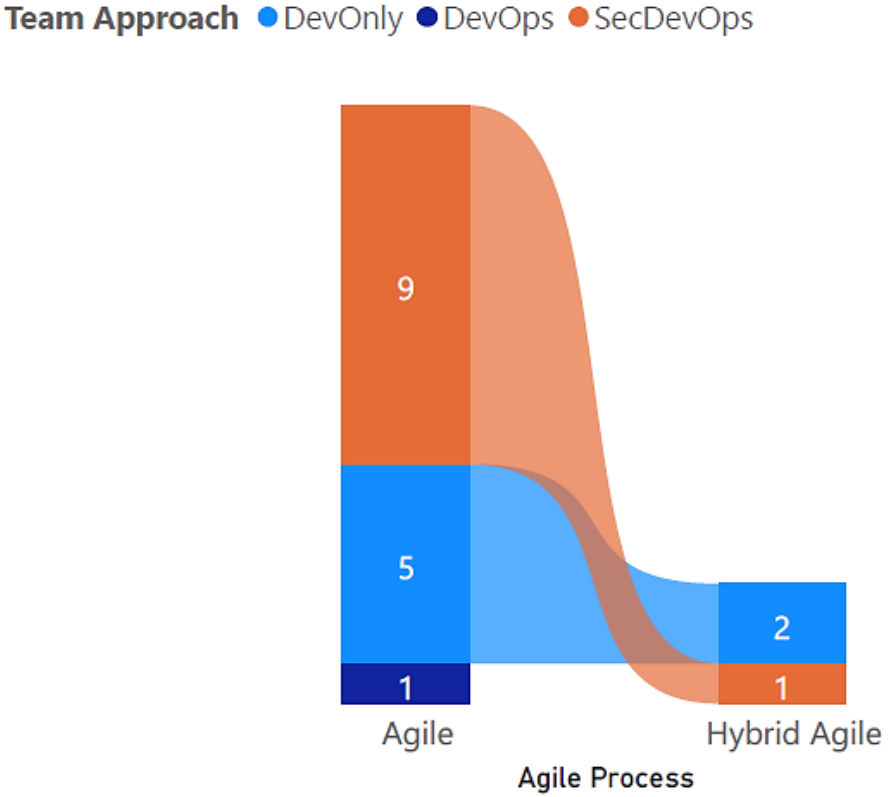


Dataset Demographics

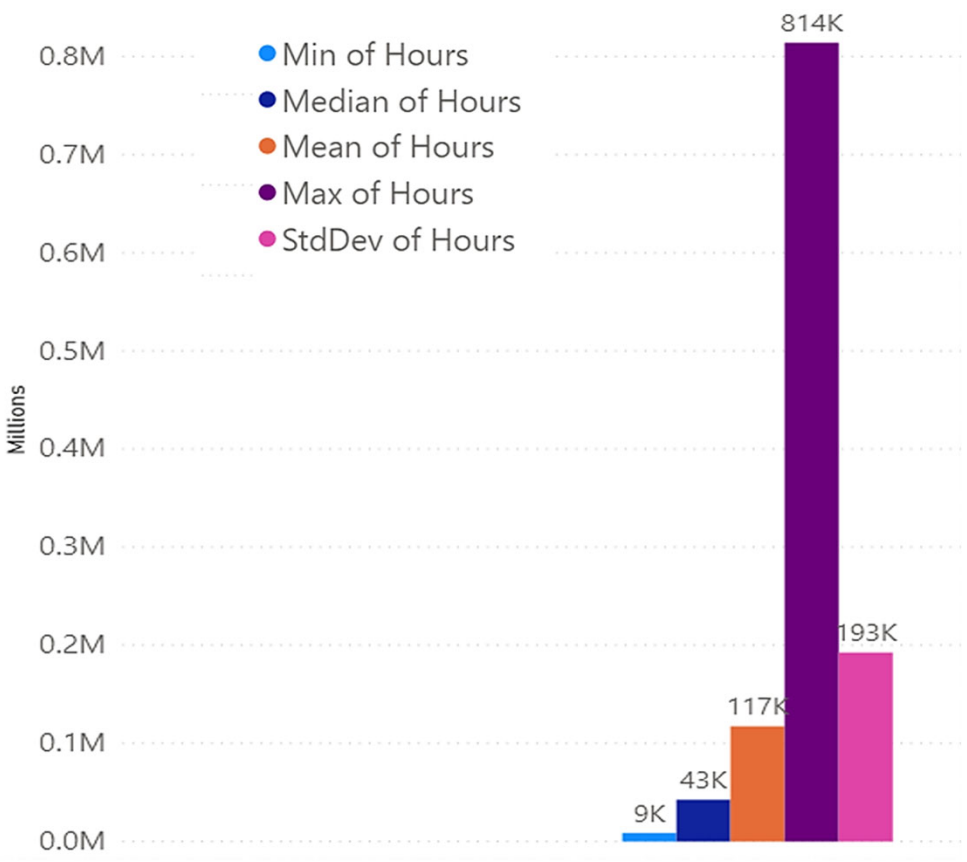
Number of Datapoints by Agile Process and Framework

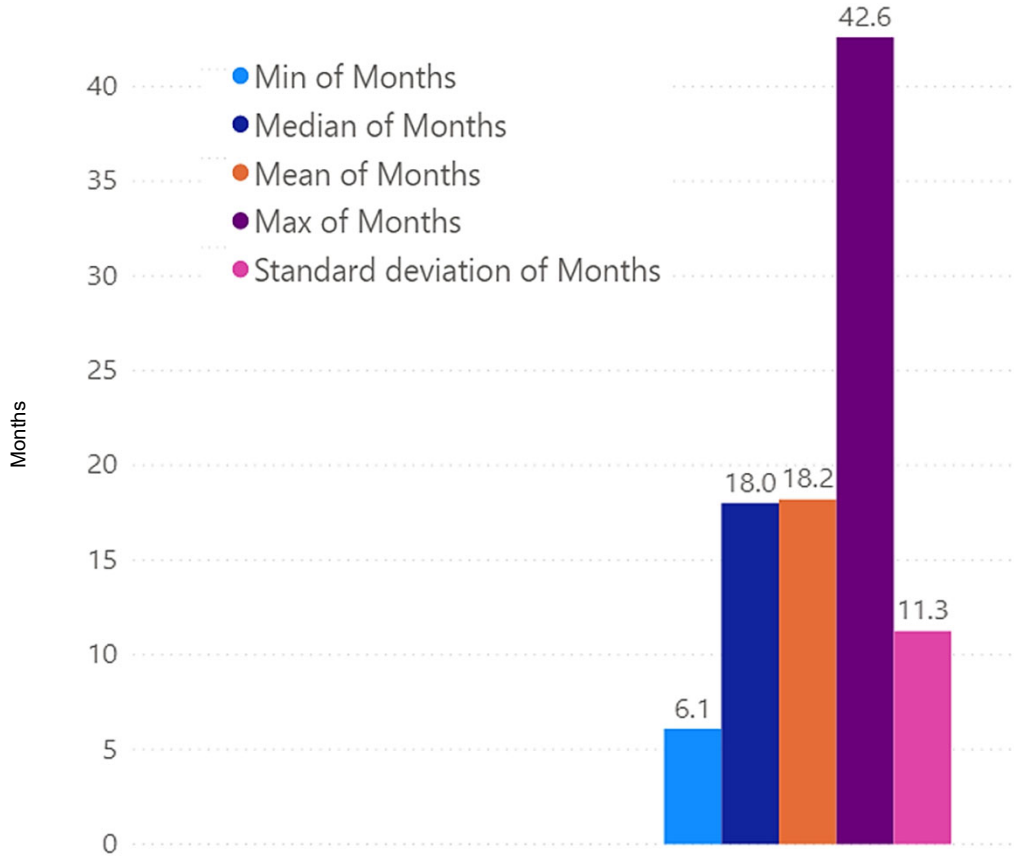


Number of Datapoints by Agile Process and Team Approach



Descriptive Statistics





Descriptive Statistics

Size Measure	Min	Median	Max	StdDev
Functional Stories	16	95	1,881	441
Issues	75	842	5,744	1,547
Stories	27	424	4,964	1,274
Story Points	602	2,708	24,492	6,109
UFP	86	659	9,628	2,368
SiFP	94	712	10,650	2,631

Relevant Range

- 20-5,000 stories
- 10-2,000 functional requirements
- 80-11,000 function points
- 9-200 Peak Staff FTEs

When selecting a regression model, consider the relevant range of each independent variable

Effort and Schedule Models

Effort Benchmarks

Category	Benchmark	25th Quartile	Median	75th Quartile	StdDev	CV
Effort	Hours/Functional Story	410	494	653	261	47%
	Hours/UFP	61	81	107	40	46%
	Hours/SiFP	57	71	100	39	47%

Practical Application:

- For example, in practice, analysts can develop an effort estimate by taking the estimated size (e.g., SiFP = 200) multiplied by the appropriate effort benchmark (median value from lookup table above):

$$Effort = Size\ x(\text{Effort Benchmark}) = 200\ x\ (71) = 14{,}200\ \text{hours}$$

Schedule Benchmarks

Category	Benchmark	25th Quartile	Median	75th Quartile	StdDev	CV
Schedule	Functional Story/FTE/ Month	0.19	0.28	0.32	0.14	47%
	UFP / FTE / Month	1.2	1.8	2.1	0.9	50%
	SiFP / FTE / Month	1.3	2.1	2.3	1.1	52%

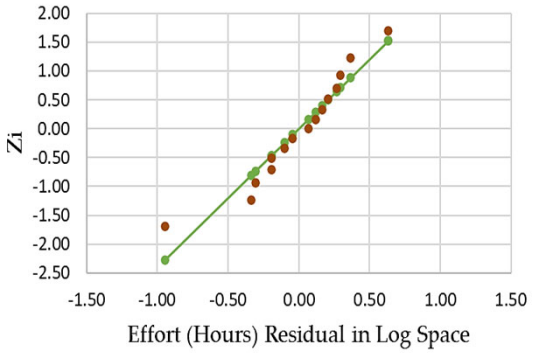
Practical Application:

- For example, an analyst can develop a schedule estimate by taking the estimated size (e.g., SiFP = 200), dividing by the appropriate schedule benchmark (median value from lookup table above), and by the estimated peak staff (e.g., FTE = 10)

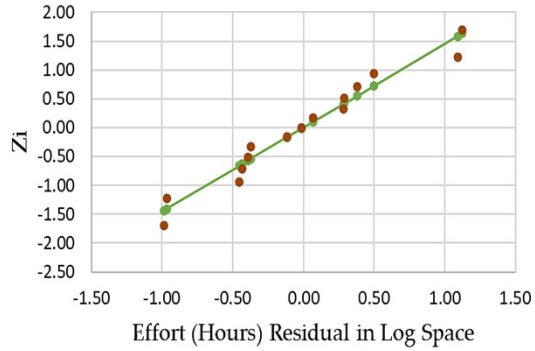
$$Schedule = Size \; \boldsymbol{x} \; (Schedule\ Benchmark)^{-1} \boldsymbol{x} \; (Peak\ FTE)^{-1} = 200 \; \boldsymbol{x} \; (2.1)^{-1} \boldsymbol{x} \; (10)^{-1} = 10 \; months$$

Effort Estimation Models

Model	CER	N	SE	R^2	R^2_{adj}	R^2_{pred}	MAD
1	$E = 935.5 x REQ^{0.882}$	15	0.39	89.6%	88.8%	85.9%	31.3%
	Effort (E) = Total final development hours REQ = Functional stories from backlog, RTM, or FRD						

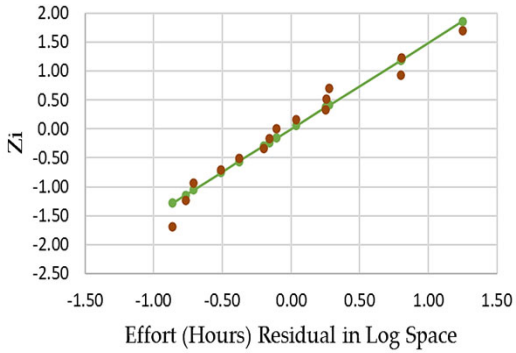


Model	CER	N	SE	R^2	R^2_{adj}	R^2_{pred}	MAD
2	$E = 1365 x STORY^{0.6228}$	15	0.66	70.2%	67.9%	59.0%	54.1%
	Effort (E) = Total final development hours STORY = Total stories obtained from JIRA backlog						

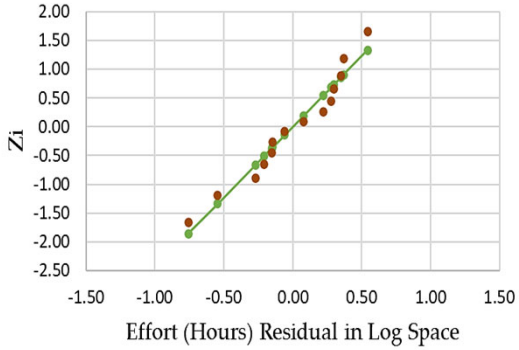


Effort Estimation Models

Model	CER	N	SE	R^2	R^2_{adj}	R^2_{pred}	MAD
3	$E = 604.3 x ISSUES^{0.6879}$	15	0.64	71.5%	69.3%	59.4%	51.6%
	Effort (E) = Total final development hours						
	ISSUES = Sum of stories, bugs, tasks, epics, or any other fixes						

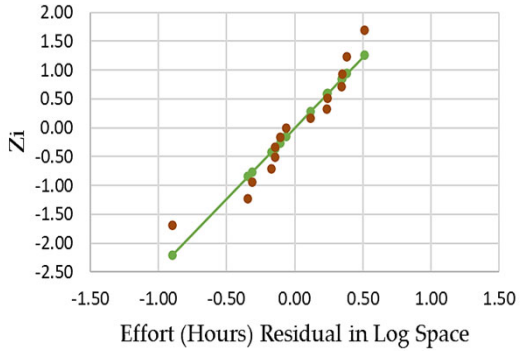


Model	CER	N	SE	R^2	R^2_{adj}	R^2_{pred}	MAD
4	$E = 206.5 x STY\_PTS^{0.6842}$	14	0.39	84.4%	83.1%	78.2%	32.7%
	Effort (E) = Total final development hours						
	STY_PTS = Story points obtained from JIRA backlog						

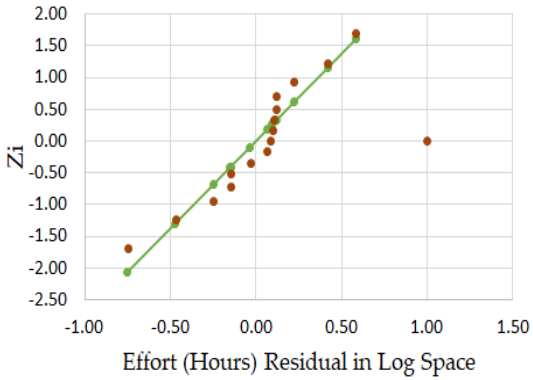


Effort Estimation Models

Model	CER	N	SE	R^2	R^2_{adj}	R^2_{pred}	MAD
5	$\mathbf{E = 189.5 x UFP^{0.8747}}$	15	0.38	90.0%	89.3%	85.6%	31.6%
	Effort (E) = Total final development hours UFP = Total unadjusted function points						

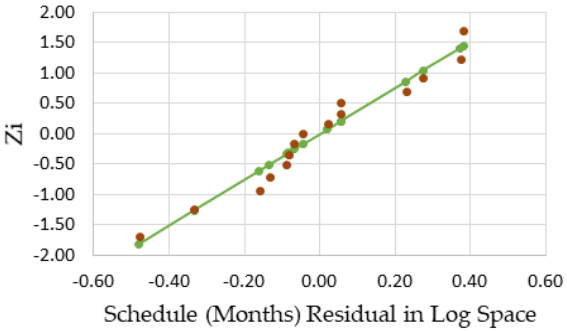


Model	CER	N	SE	R^2	R^2_{adj}	R^2_{pred}	MAD
6	$\mathbf{E = 261.1 x SiFP^{0.7708} x 1.615^{D1}}$	15	0.35	92.0%	90.7%	86.5%	25.9%
	Effort (E) = Total final development hours SiFP= Simple Function Point D1= Dummy variable (scope), where full development =1 and enhancement =0						

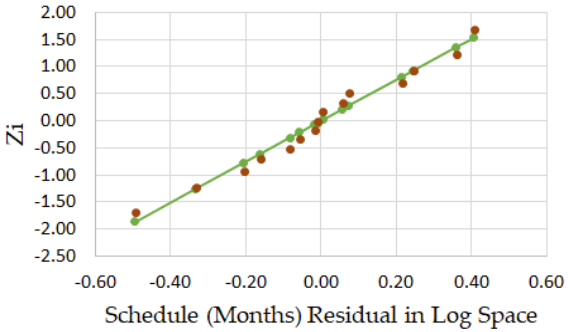


Schedule Estimation Models

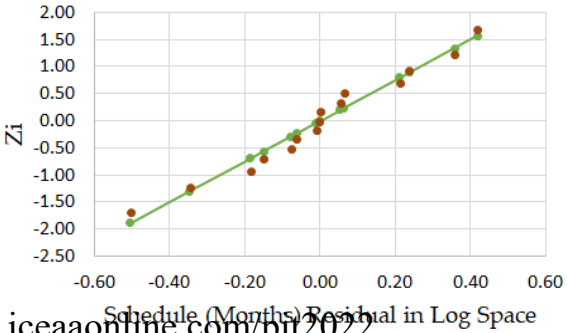
Model	SER	N	SE	R^2	R^2_{adj}	R^2_{pred}	MAD
1	$S = 2.685 x \text{REQ}^{0.2135} x 2.718^{D1}$	15	0.26	88.4%	86.5%	81.8%	18.8%
	Schedule (S)= Total final development months REQ= Functional stories from backlog, RTM, or FRD D1= Dummy variable (scope), where full development =1 and enhancement =0						



Model	SER	N	SE	R^2	R^2_{adj}	R^2_{pred}	MAD
2	$S = 1.938 x \text{UFP}^{0.2025} x 2.739^{D1}$	15	0.27	88.1%	86.2%	80.8%	18.6%
	Schedule (S)= Total final development months UFP= Total unadjusted function points D1= Dummy variable (scope), where full development =1 and enhancement =0						



Model	SER	N	SE	R^2	R^2_{adj}	R^2_{pred}	MAD
3	$S = 2.009 x \text{SiFP}^{0.1923} x 2.826^{D1}$	15	0.27	88.0%	86.0%	80.4%	18.4%
	Schedule (S)= Total final development months SiFP= Simple Function Point D1= Dummy variable (scope), where full development =1 and enhancement =0						



How do the Regression Models Rank?

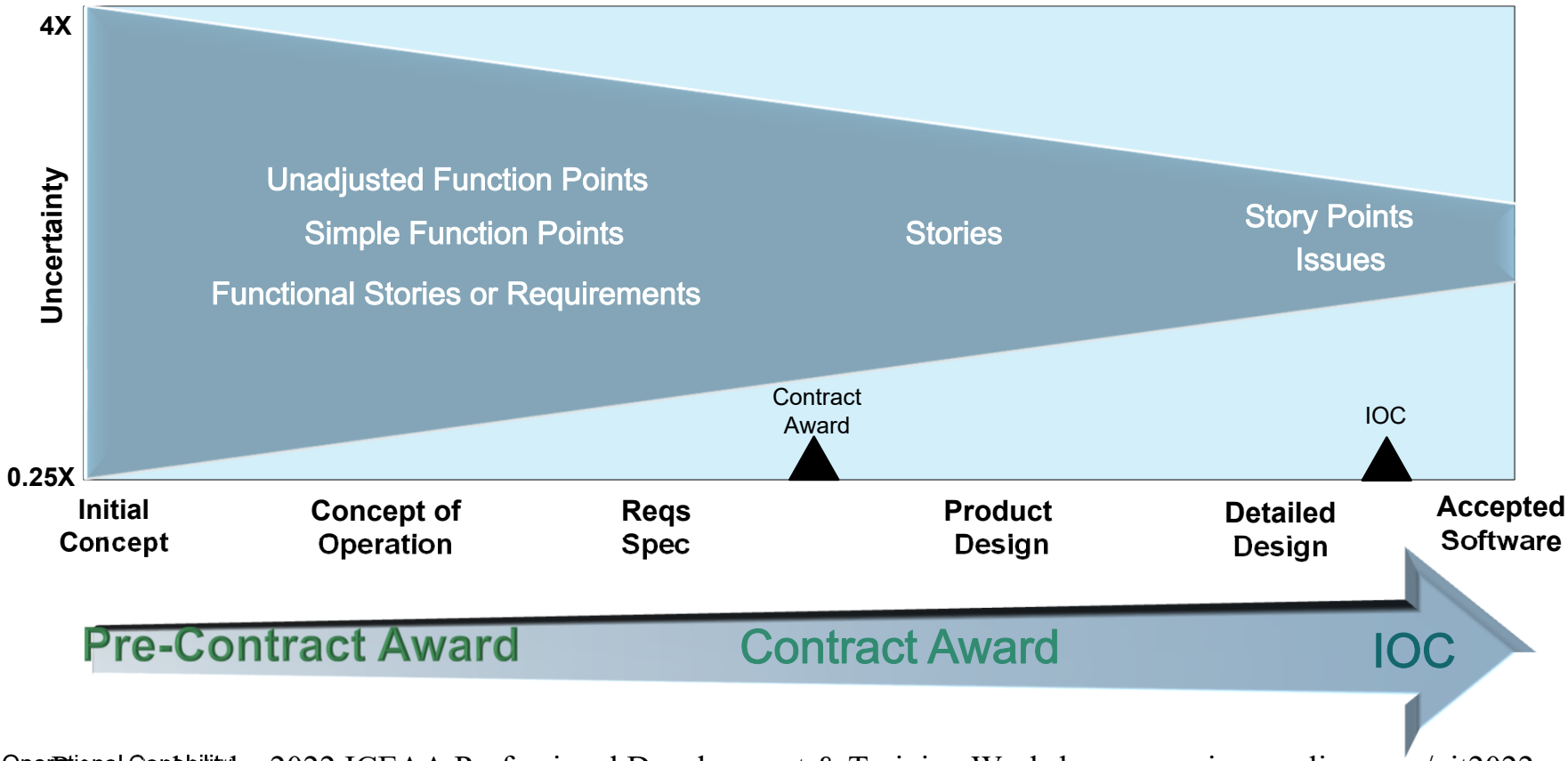
Rank	Model Equation	R^2_{adj}	R^2_{pred}	MAD
Effort Models				
1	$E = 261.1\ x\ SiFP^{0.7708}\ x\ 1.615^{D1}$	90.7%	86.5%	25.9%
2	$E = 189.5\ x\ UFP^{0.8747}$	89.3%	85.6%	31.6%
3	$E = 935.5\ x\ REQ^{0.882}$	88.8%	85.9%	31.3%
4	$E = 206.5\ x\ STY\_PTS^{0.6842}$	83.1%	78.1%	32.7%
5	$E = 604.3\ x\ ISSUES^{0.6879}$	69.3%	59.4%	51.6%
6	$E = 1365\ x\ STORY^{0.6228}$	67.9%	59.0%	54.1%
Schedule Models				
1	$S = 2.685\ x\ REQ^{0.2135} x\ 2.718^{D1}$	86.5%	81.8%	18.8%
2	$S = 1.938\ x\ UFP^{0.2025} x\ 2.739^{D1}$	86.2%	80.8%	18.6%
3	$S = 2.009\ x\ SiFP^{0.1923} x\ 2.826^{D1}$	86.0%	80.4%	18.4%

Simple Function Point (SiFP), Unadjusted Function Points (UFP), and Functional Stories (REQ) are stronger predicters to both effort and schedule for agile projects

Results

When to Use the Models

When choosing the appropriate estimation model, analysts should consider:
1 program's lifecycle maturity, and 2 which size measures are available at that time



IOC = Initial Operational Capability

Model Limitations

Internal Threats

- Dataset timeframe (2014-2021) raises potential issues as the earlier projects (2014, 2016, 2018) may have used agile processes tailored to fit the agency's need.

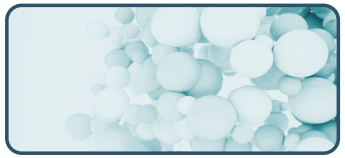
External Threats

- Models proved to be effective for estimating agile in the DHS context. However, we cannot generalize beyond this group.
- Agencies may not have access to Backlog, FRD or RTM for SiFP analysis

Constructive Threats

- Small dataset does not allow for detecting effect with greater power (Overfitting)
- Need a larger dataset to draw more confident statistical conclusions

Main Takeaways



Wider range of **sizing measures (6)** to estimate future agile software programs and evaluate contractor proposals

SiFP and UFP proved to be the most accurate predictors of agile software development effort and schedule

Analysis reveals that "**Functional Story**" is an effective predictor of effort and schedule, and easy to obtain

Popular agile measures such as story points, stories, and issues *are not as effective predictors of effort and schedule*

SiFP and UFP can be calculated early in the program **allowing estimation** from contract proposal through IOC (when popular agile measures are difficult to obtain)

Acknowledgement

Name	Organization	Role
Katie Geier-Noreiga	DHS CAD	Sponsor
Tommy Reynolds	DHS CAD	Sponsor
Blaze Smallwood	Booz Allen	Contributor
Carol Dekkers	Galorath	Contributor
Ryan Kapsak	Galorath	Contributor
Jonathan Greene	Galorath	Contributor
Katharine Mann	DHS CAD	Contributor
Ryan Hoang	DHS CAD	Contributor
Jose Torres	DHS CAD	Reviewer
Dr. Christian Smart	Galorath	Reviewer
David DeWitt	Galorath	Reviewer
Duncan Thomas	Galorath	Reviewer

Presenter Information

Dr Wilson Rosa
Assistant Director, DHS Cost Analysis Division, OCFO
wilson.rosa@hq.dhs.gov

Sara Jardine
Senior Cost Analyst, DHS Cost Analysis Division, OCFO
sara.jardine@associates.hq.dhs.gov
sjardine@galorath.com

Chad Lucas
Senior Cost Analyst, DHS Cost Analysis Division, OCFO
chad.lucas@associates.hq.dhs.gov
clucas@galorath.com