29th Annual INCOSE international symposium

Orlando, FL, USA
July 20 - 25, 2019

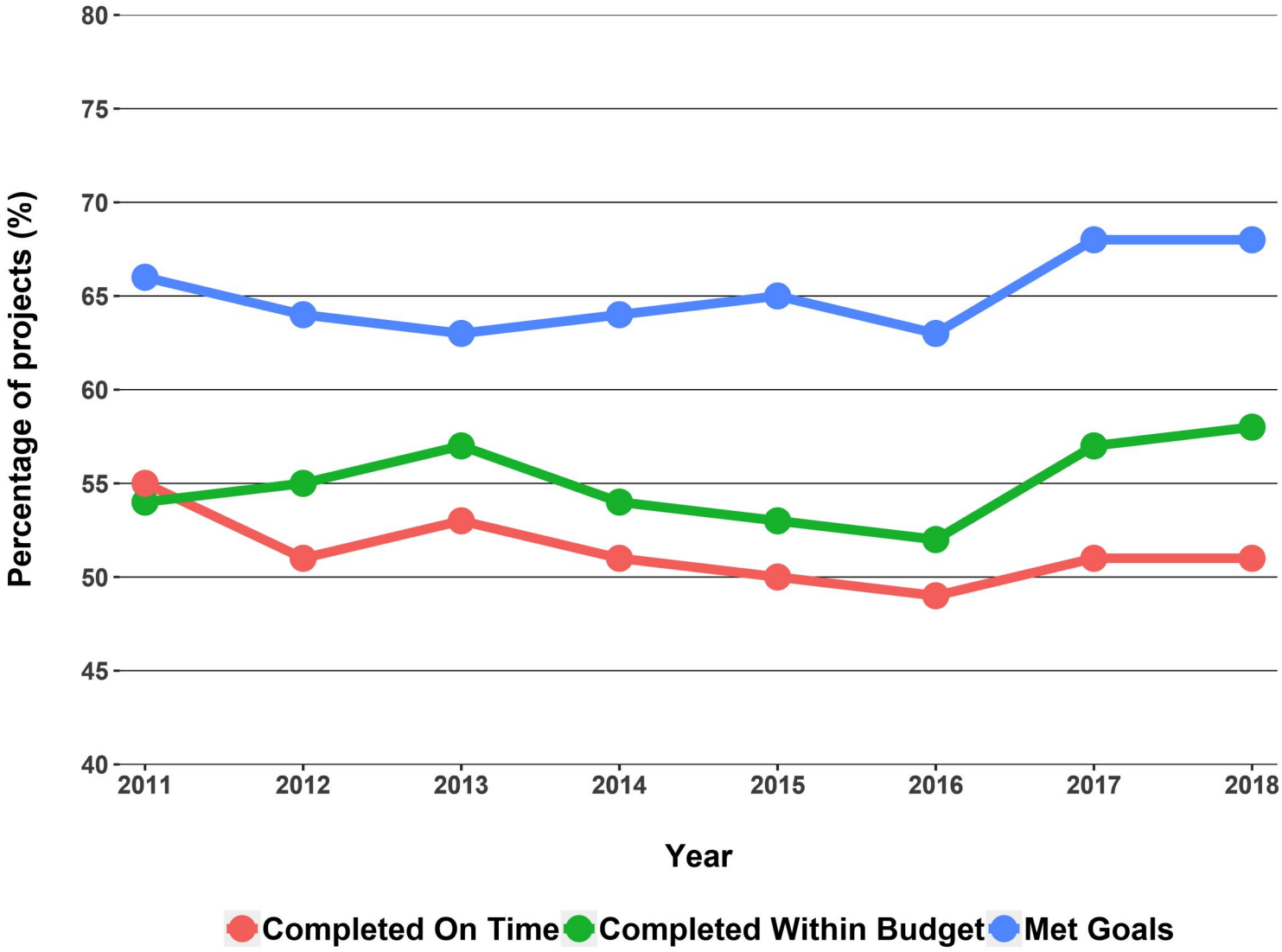
Can We Use Wisdom-of-the-Crowd to Assess Risk of Systems Engineering Failures?

Georgios Georgalis and Dr. Karen Marais

This research work was partially funded by the Systems Engineering Research Center (SERC)

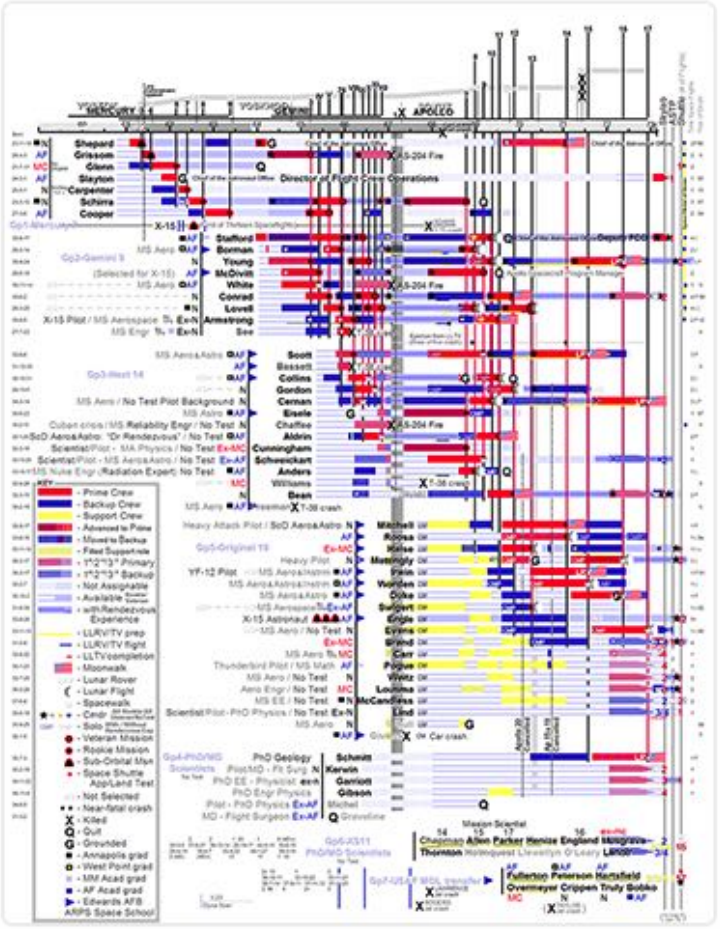
SYSTEMS ENGINEERING RESEARCH CENTER

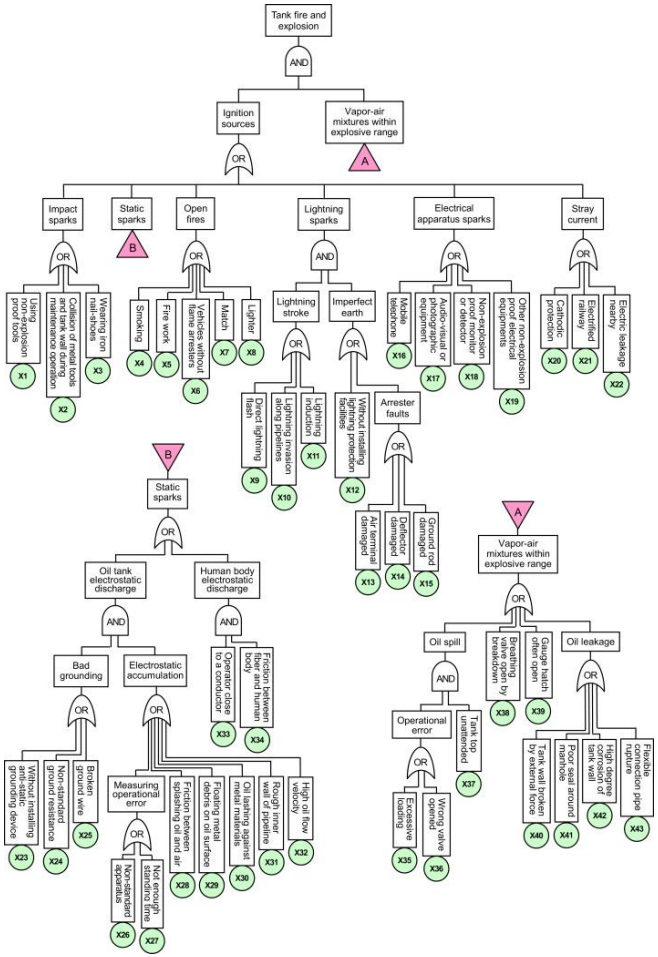
Projects continue to fail despite new practices in systems engineering and project management



Data from Project Management Institute (PMI): Pulse of the Profession 2018

Systems engineers use various methods to assess risk but timely assessment is still a challenge





Our overall goal: Can Wisdom-of-the-Crowd (WoC) improve how we do risk assessment?

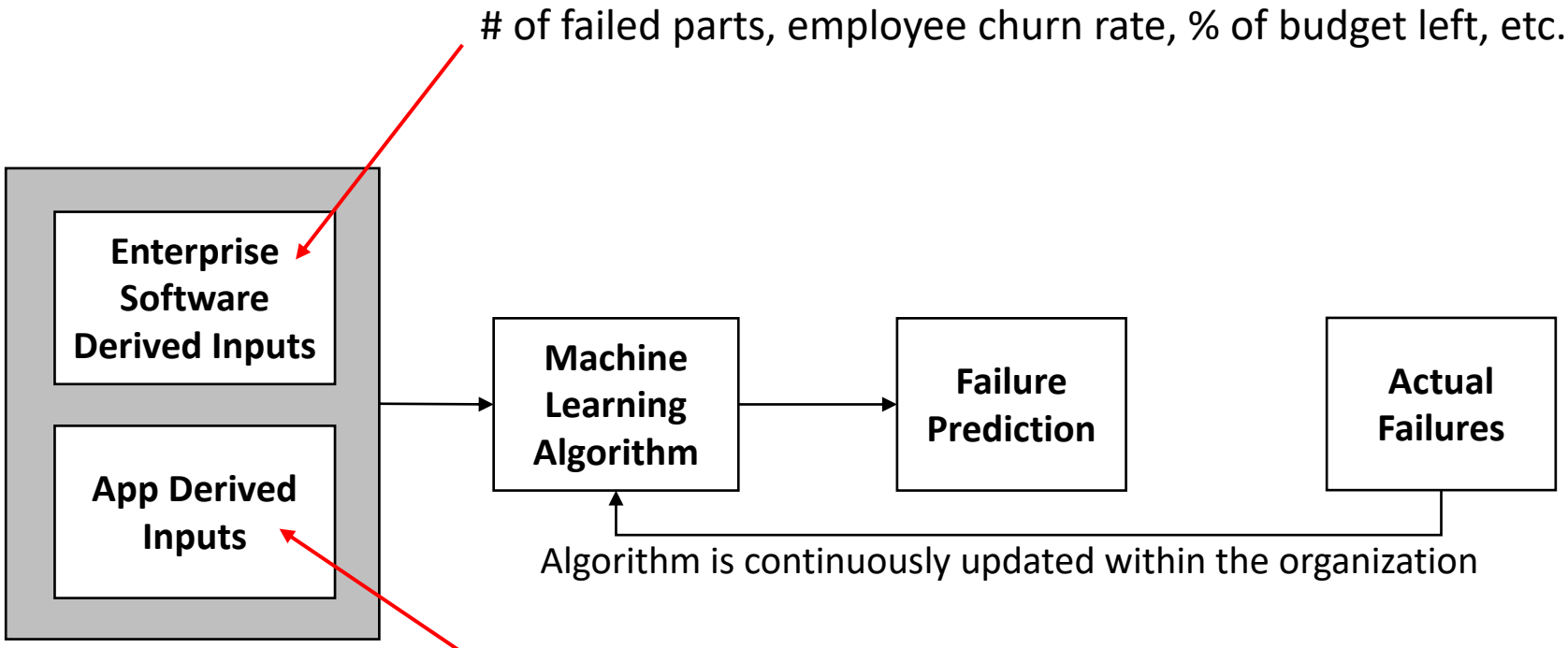
Basic principle of WoC:

- A large group of people can be more accurate in their estimates than a single subject-matter expert
- First experiments included quantitative estimates

WoC in risk assessment:

- Use WoC indicators to augment traditional risk information about a project
- WoC indicators come from employees answering questions via an app
- Goal: To predict future systems engineering failures

Our overall goal: Can Wisdom-of-the-Crowd (WoC) improve how we do risk assessment?



- Team performance
- Cognitive biases
- Safety practices
- Risk perception
- Indirect actions or habits (e.g., # of project outputs)
- Estimates about project risk (e.g., schedule)
- Personality
- Critical Success Factors

Our goal for this paper: Are crowd estimates about project risk more accurate than individual estimates?

- Before we apply WoC-based risk assessment in industry, we want to see if it works at all

- Student projects: Can students give us accurate risk estimates about their project?

- If WoC applies, we expect risk estimates of a crowd to be closer to the truth than any single individual's

Three steps to answer whether WoC applies in risk assessment

- Collect qualitative estimates of three project metrics
  - Budget
  - Schedule
  - Technical requirements

- Find how accurate these estimates are

- Compare two cases to confirm whether WoC applies:

1. Consider the estimates of the individuals separately

2. Consider the estimates grouped as a “crowd”

We collected data from 18 student design teams from two senior-level courses over two semesters

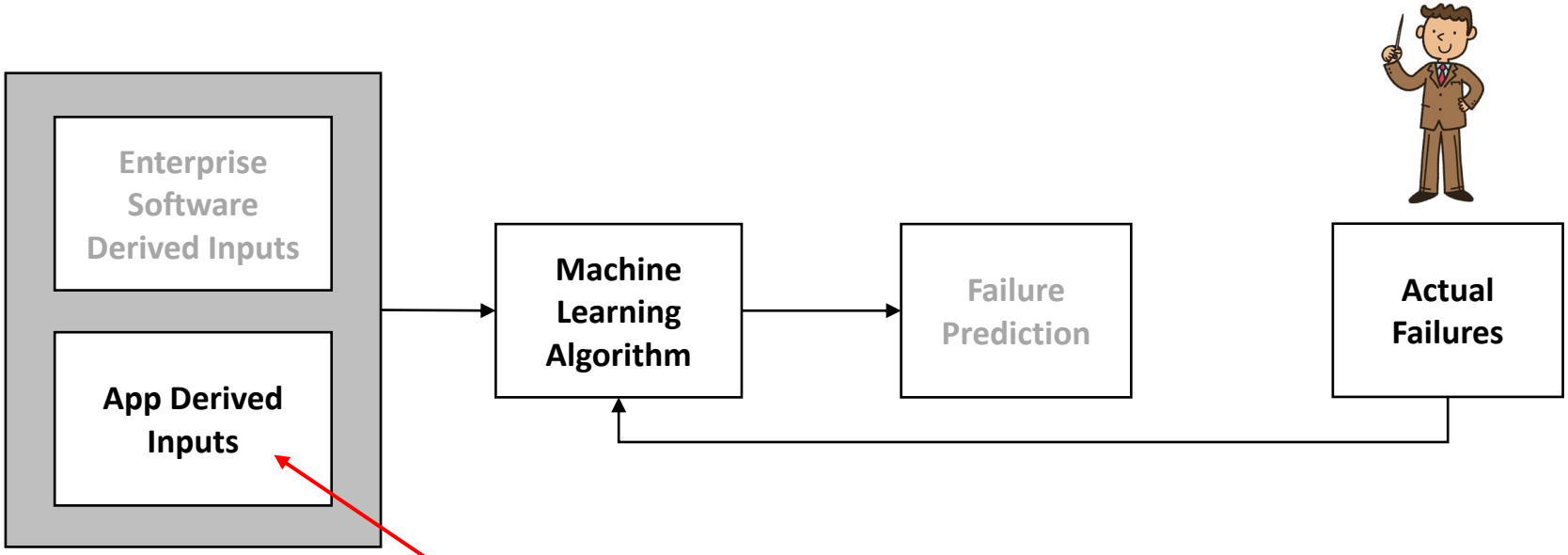
Instructors give us the “true” qualitative metrics of each project (required to check accuracy of student estimates)

Students answer three out of the nine question categories shown earlier

- Estimates about project risk
- Cognitive biases
- Safety Practices

- We used a Qualtrics survey to collect the data every week

Can students on a project give us accurate estimates of what is currently happening with the project?



- Team performance
- **Cognitive biases ("Individual actions & decisions") - 9**
- **Safety practices ("Team Actions and Archetypes") - 4**
- Risk perception
- Indirect actions or habits (e.g., # of project outputs)
- **Estimates about project risk (e.g., schedule) - 3**
- Personality
- Critical Success Factors

We asked the instructors three questions, one for each project metric

What is currently true about the project spending, compared to what you initially planned?

	Project spending
Project 1	On budget
Project 2	On budget
Project 3	On budget
Project 4	On budget
Project 5	On budget

Cli ns...

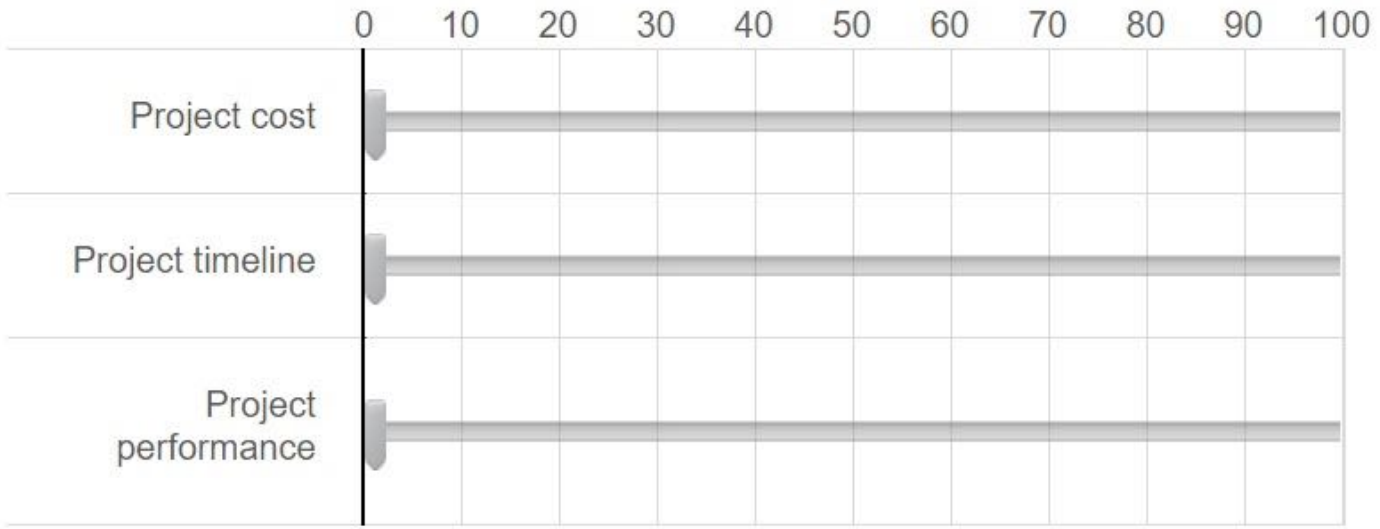
Under budget
On budget
Over budget

Students respond to the 16 questions on a website/ smartphone-accessible Qualtrics survey

Which of the following reflects your *current* estimate about your project's timeline?

- We are running behind schedule.
- We are running on schedule.
- We are running ahead of schedule.

How confident are you in your estimates?



9 yes/no questions capture individual actions that may relate to poor judgment

- Bandwagon effect: Tendency to do or believe what others do or believe

Do you suffer from bandwagon effect?

No, of course not!

- Focusing effect: Tendency to place too much importance on one aspect

Do you suffer from focusing effect?

- When possible, phrase questions as hard to game and in context of a student project

Bandwagon effect

Focusing effect

4 yes/no questions capture collective team behaviors that may relate to poor safety practice

Fixing symptoms rather than root causes:

Using symptomatic solutions that become less effective over time resulting in problem resurfacing

We want to compare the accuracy of the estimates in two cases: “individual” and “crowd”

“Individual” case

Project Metrics

Individual Student Estimates

Instructor Estimates

Y = 1 (estimates matched)

Y = 0 (estimates did not match)

Individual Student Confidence

Individual Actions & Decisions

Individual Case

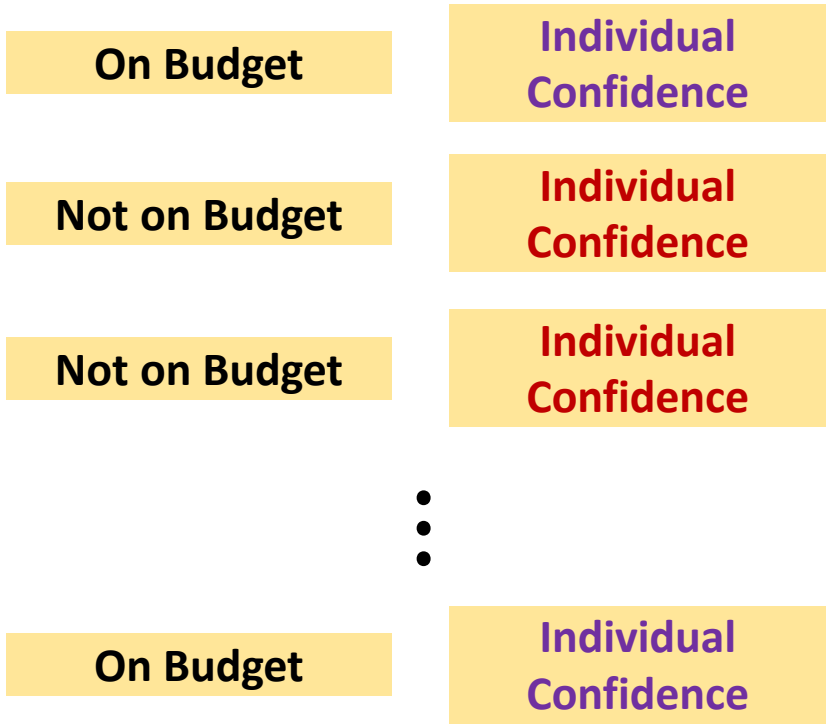
On Budget	**Individual Confidence**
Not on Budget	**Individual Confidence**
Not on Budget	**Individual Confidence**
⋮	
On Budget	**Individual Confidence**

“Crowd” case

1. Average confidence for “on budget” responses
2. Average confidence for “not on budget” responses
3. The estimate that has higher average confidence becomes the WoC estimate

On Budget → **WoC Estimate**

WoC confidence is then calculated using weighted averages



$$\text{WoC Confidence} = \frac{\sum_{i=1}^{n_1} confidence\_win + \sum_{i=1}^{n_2} (100 - confidence\_lose)}{n_1 + n_2}$$

“Crowd” case

Project Metrics

WoC Estimates

Instructor Estimates

Y = 1 (estimates matched)

Y = 0 (estimates did not match)

WoC Confidence

Team Actions & Archetypes

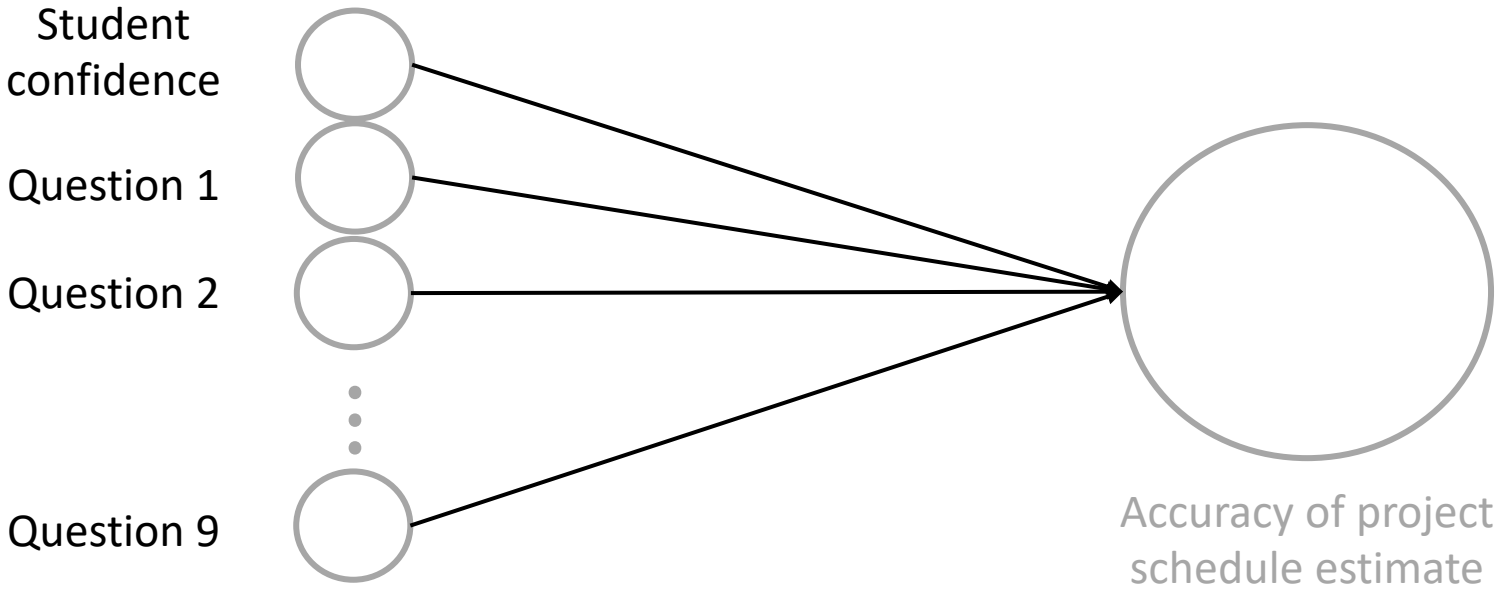
We end up with 6 models in total: An individual and crowd model for each of the 3 metrics

- Probabilistic links between the accuracy of an estimate given the inputs: confidence and answers to additional questions

- We used *mixed effects* in our models: a technique to remedy some regression assumptions

- Is $\hat{p}_{crowd} > \hat{p}_{indiv.}$?

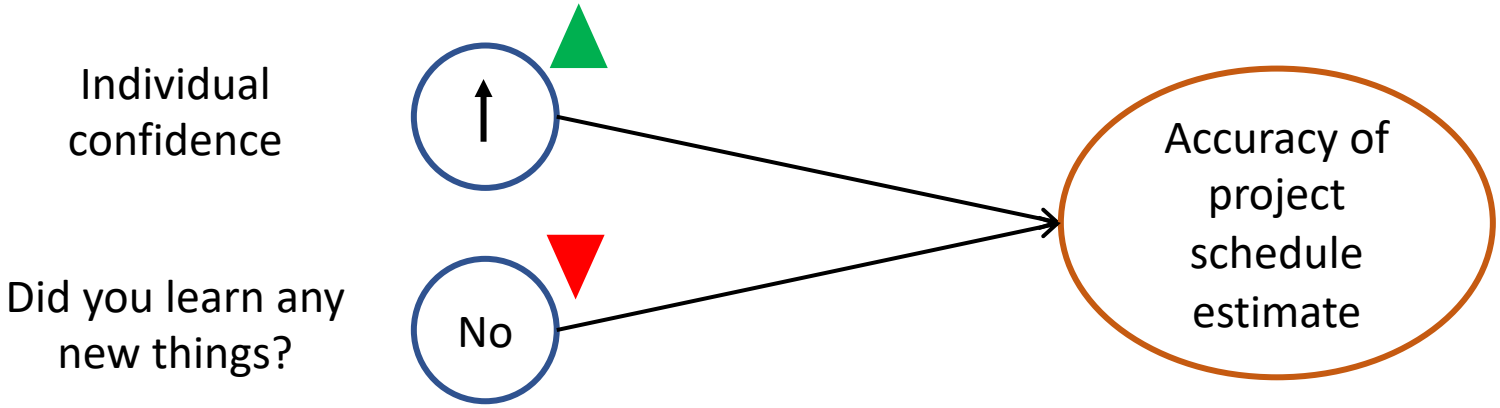
Each link comes with a p-value



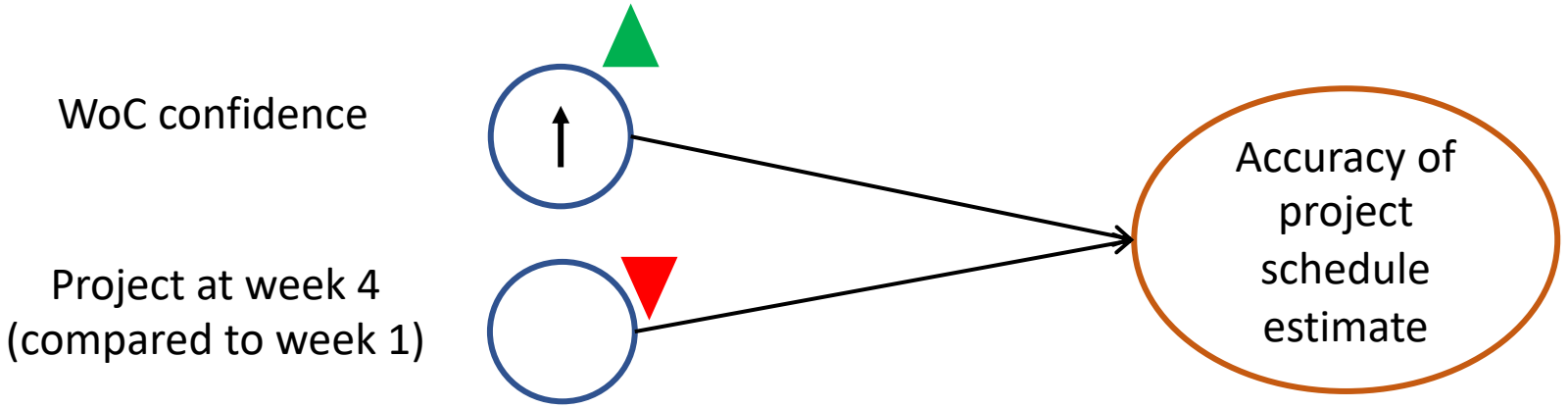
- Small p-value means there is a lot of evidence in the data and we can be confident about the correctness of the link

Both models show that increased confidence improves the accuracy of the given estimate

Individual schedule model



Crowd schedule model



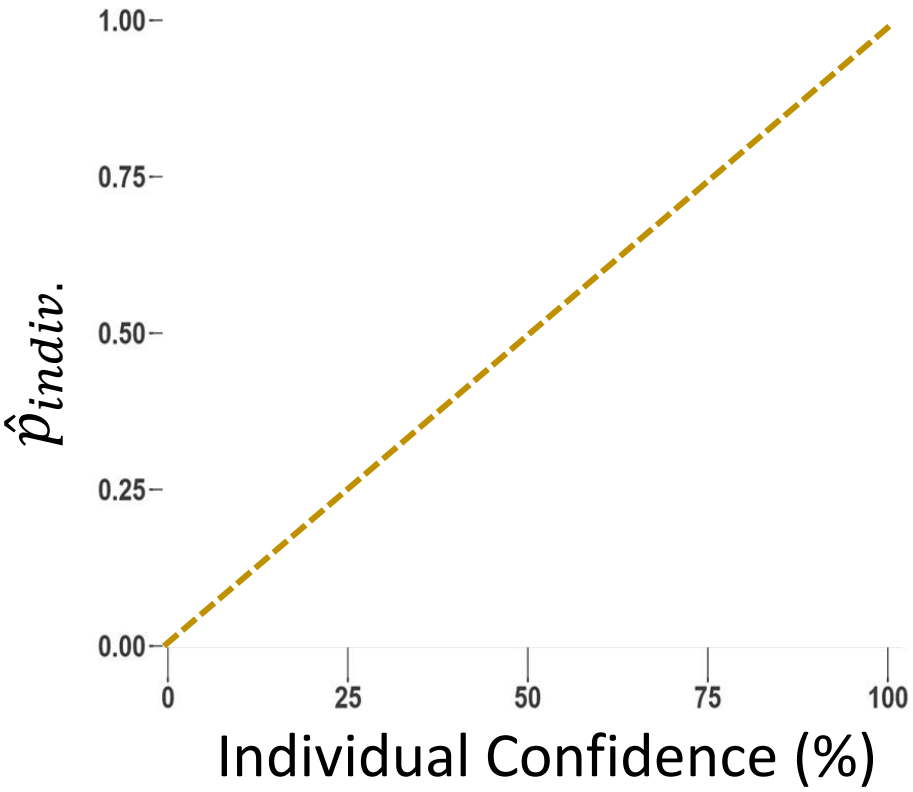
- Interpreting the logistic regression models is difficult because the correlations are expressed in log odds scales

- We can compare the probabilities of correct estimates using the marginal effect of confidence
  - Previous research confirms that higher confidence results in better estimates of systems engineers (Nolan et. al, 2018)

- How does the probability of a correct estimate vary, on average, given the confidence metric?

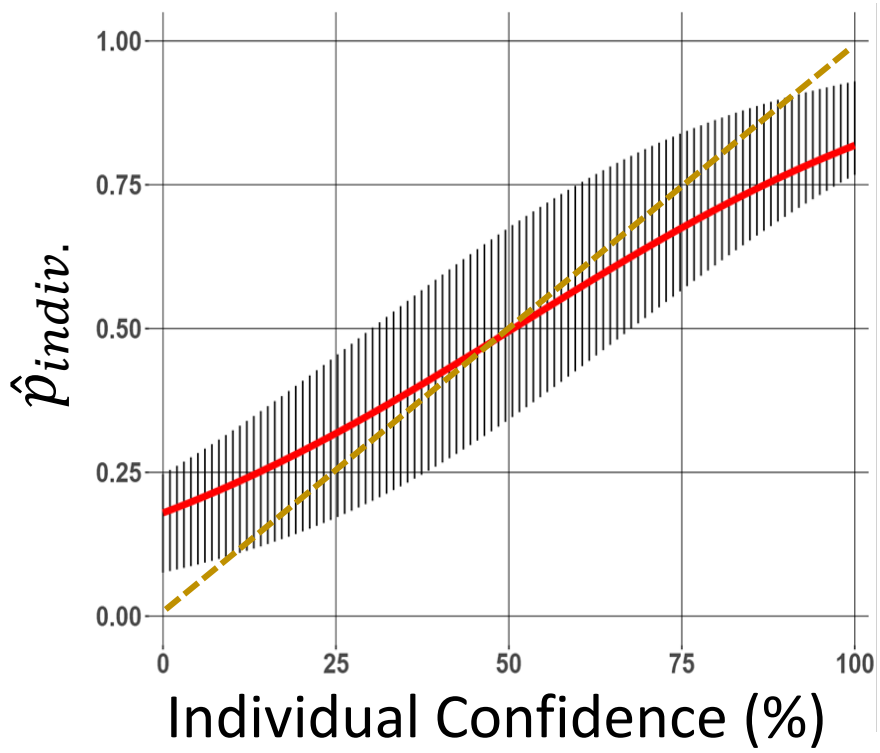
Nolan, A., Pickard, A.C., Nolan, J., Beasley, R. and Pruitt, T.C., 2018, July. How Many Systems Engineers Does It Take To Change a Light Bulb?. In *INCOSE International Symposium* (Vol. 28, No. 1, pp. 777-790).

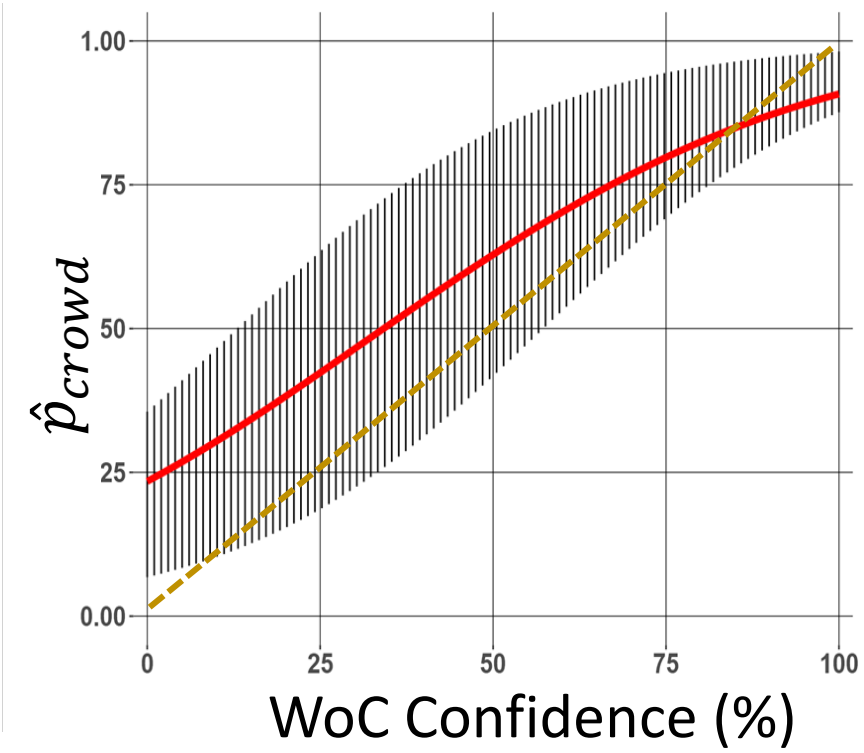
We compare the probability that a qualitative estimate is accurate in the two cases, given a confidence metric



- For confidence more than 60%, the WoC approach is more likely to provide an accurate schedule estimate compared to the individual approach

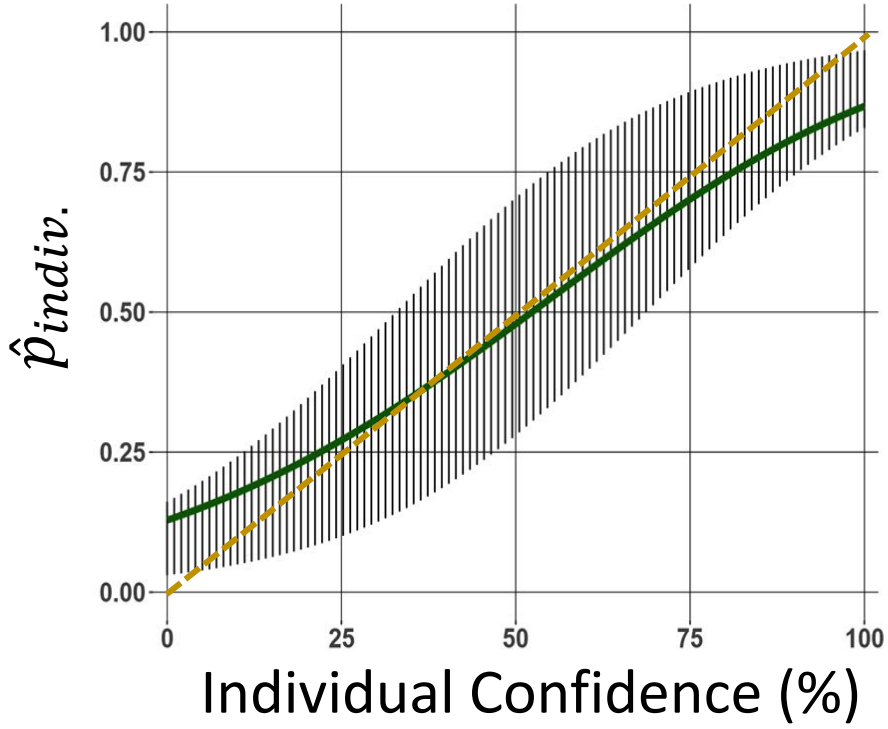
When estimating budget, the WoC approach provides more accurate estimates

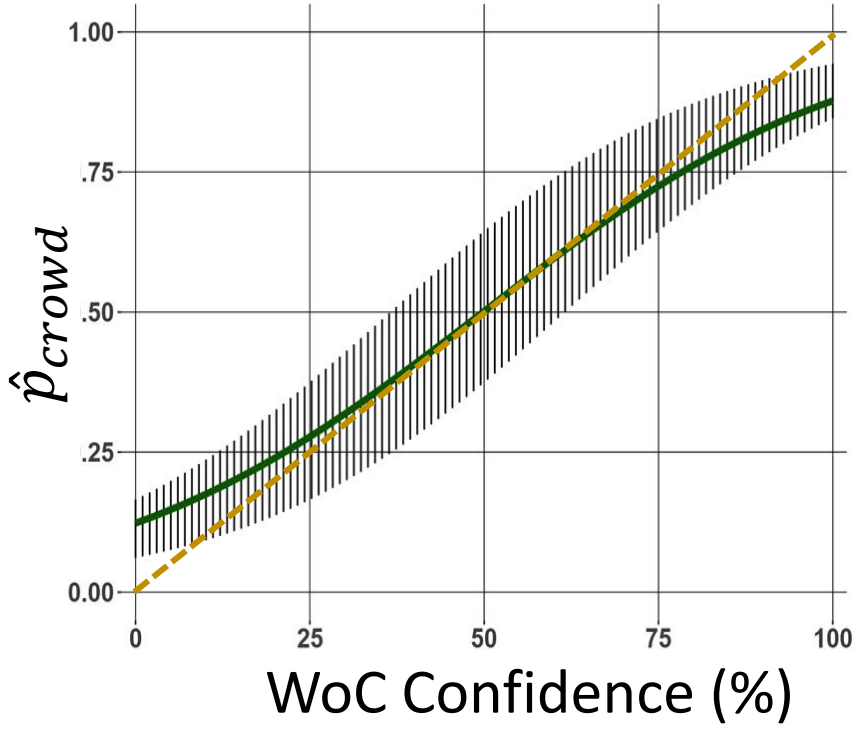




- The distribution of predicted probabilities has larger variance for budget estimation (perhaps because students do not have specific budget targets, so their estimates vary more per person/project)

When estimating technical performance, the WoC approach is better for confidence levels >50%





- A benefit of the WoC approach here is the smaller variance on the estimated probability

Summary

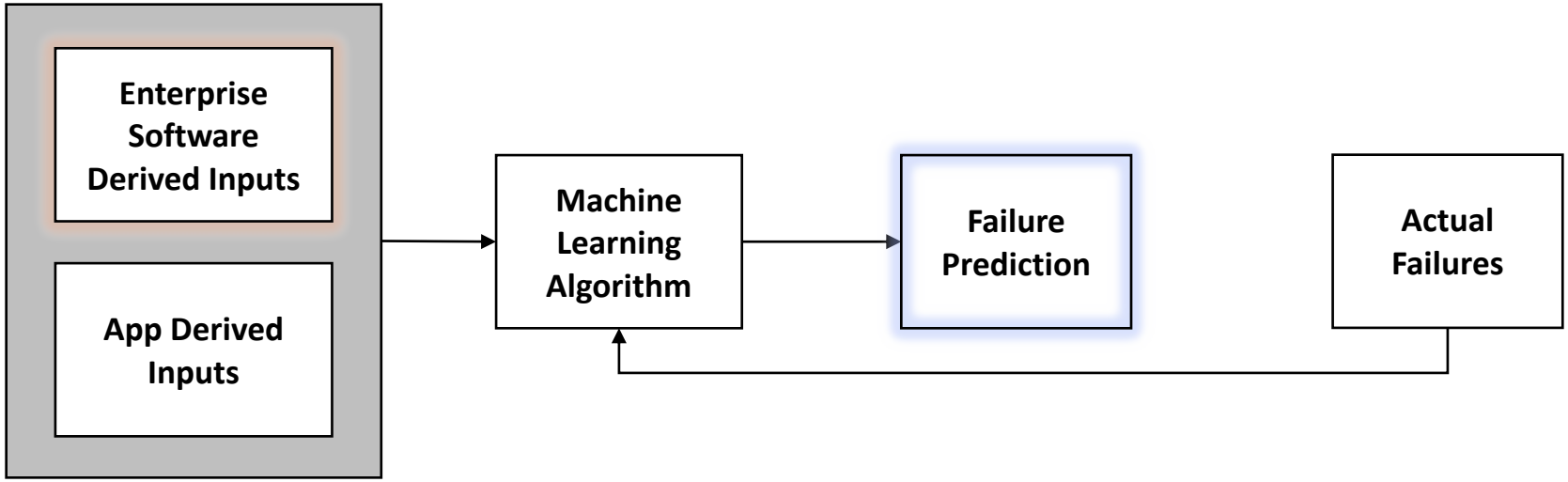
- When estimating budget, we found that the WoC approach provides more accurate estimates for ***all*** confidence values

- When estimating schedule or requirements, we found that the WoC approach provides more accurate estimates if they come with high collective confidence

- The likelihood of students providing correct estimates about project risk increases as they become more confident

Future Work

- Expand the survey to include all the categories of questions mentioned earlier, to capture more risk-related information
- Can we predict failures based on this information from the project team?
- Does the approach improve further with traditional risk inputs from the enterprise software? (industrial application)



BACKUP SLIDES

Mixed effects model relax some of the logistic regression assumptions

- Panel data: observation independence assumption does not hold
- Common approach: Include random effects for subjects (Harrison et al., 2018)
- Allow for each student to have a different intercept term

$$\hat{Y}_{j,t+1} = \log\left(\frac{\hat{p}_{j,t+1}}{1-\hat{p}_{j,t+1}}\right) = a + bX_{i,t}{}^{T} + c_i + \varepsilon_{it}$$

- $c_{\mathrm{i}} \sim N(0, \sigma_i{}^2)$, $\varepsilon_{it} \sim N(0, \sigma_\varepsilon{}^2)$
- lme4 package for R for mixed effects modeling, built by Prof. Douglas Bates

Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E., Robinson, B.S., Hodgson, D.J. and Inger, R., 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *6*, p.e4794.