We are collaborating with other researchers to apply causal learning to learn how to control costs in software development and sustainment.

DoD Problem
• DoD leadership needs to understand why software costs so much.
• DoD program offices need to know where to intervene to control software costs.

Why Causal Learning?
To reduce costs, the causes of an outcome (good or bad) need to be considered. Correlations are insufficient due to Simpson's Paradox. For example, in the figure below, if you did not segment your data by team (User Interface [UI] and Database [DB]), you might conclude that increasing domain experience reduces code quality (downward line); however, within each team, it’s clear that the opposite is true (two upward lines). Causal learning identifies when factors such as team membership explain away (or mediate) correlations, and it works for much more complicated datasets too.

Our Solution
Working with collaborators, we will jointly apply causal learning to their datasets to establish key cause-effect relationships among project factors and outcomes.

Our collaborators include the University of Southern California, U.S. Army, and a static code analysis tool vendor.

For example, for effort, we might have this causal graph:

This graph tells us that increasing stakeholder reviews (SR) and domain experience (DE) improves the effectiveness of requirements, analysis, coding, and testing, thereby improving quality.

If the dataset is proprietary, the SEI trains the collaborator to perform causal searches on their own. The SEI then needs information only about what dataset and search parameters were used as well as the resulting causal graph.

Summary
Causal models offer better insight for program control than models based on correlation. Knowing which factors drive which program outcomes is essential to sustain the warfighter by providing higher quality, secure software in a timely and affordable manner.