Doing Empirical Software Engineering

Defect Prediction

Elaine Weyuker

Tom Ostrand

AT&T Labs – Research

July 2009

RESEARCH GOAL

To determine which files of a large industrial software system with multiple releases are particularly likely to contain large numbers of defects in the next release.

A High-Level Prescription

- Develop a theory: there are file properties that can be determined and used to identify defect-prone files.
- Get initial interest from practitioners – this is a form of a priori validation that this is a "real" problem.
- Find an initial system to study.
- Do preliminary empirical studies to identify these properties.
- Present preliminary results to practitioners and **listen** to feedback.
- Build statistical models.
- Collaborate with researchers with complimentary skills
- Perform multiple case studies using real large systems to validate results and determine generality.
- Present results to both practitioners and researchers and **listen** to feedback.
- Recognize that the process may be very time-consuming.
- Automate!! No one will use your results if it requires a large amount of time and expertise to use your technology.

ISSUES TO CONSIDER

- Awareness by practitioners
- Relevant to practitioners
- Useable by practitioners