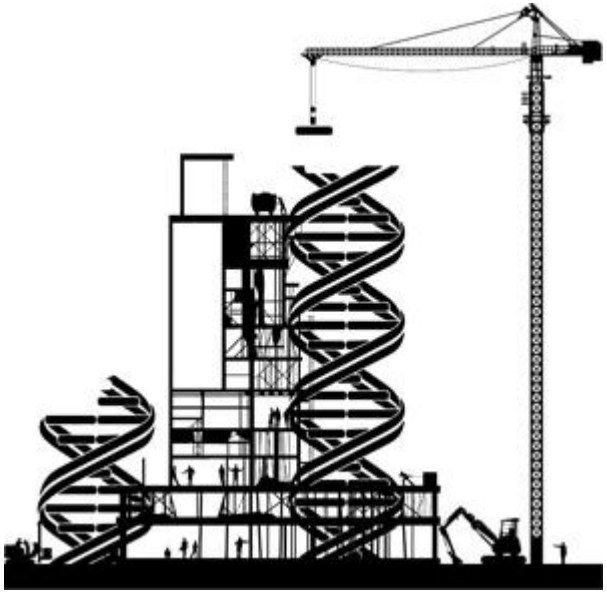
SATE V Synthetic



Method and Results

NIST

March 14, 2014

Aurelien Delaitre

SATE V Organizer

aure at nist dot gov

Disclaimer

Certain instruments, software, materials, and organizations are identified in this paper to specify the exposition adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the instruments, software, or materials are necessarily the best available for the purpose.

Input

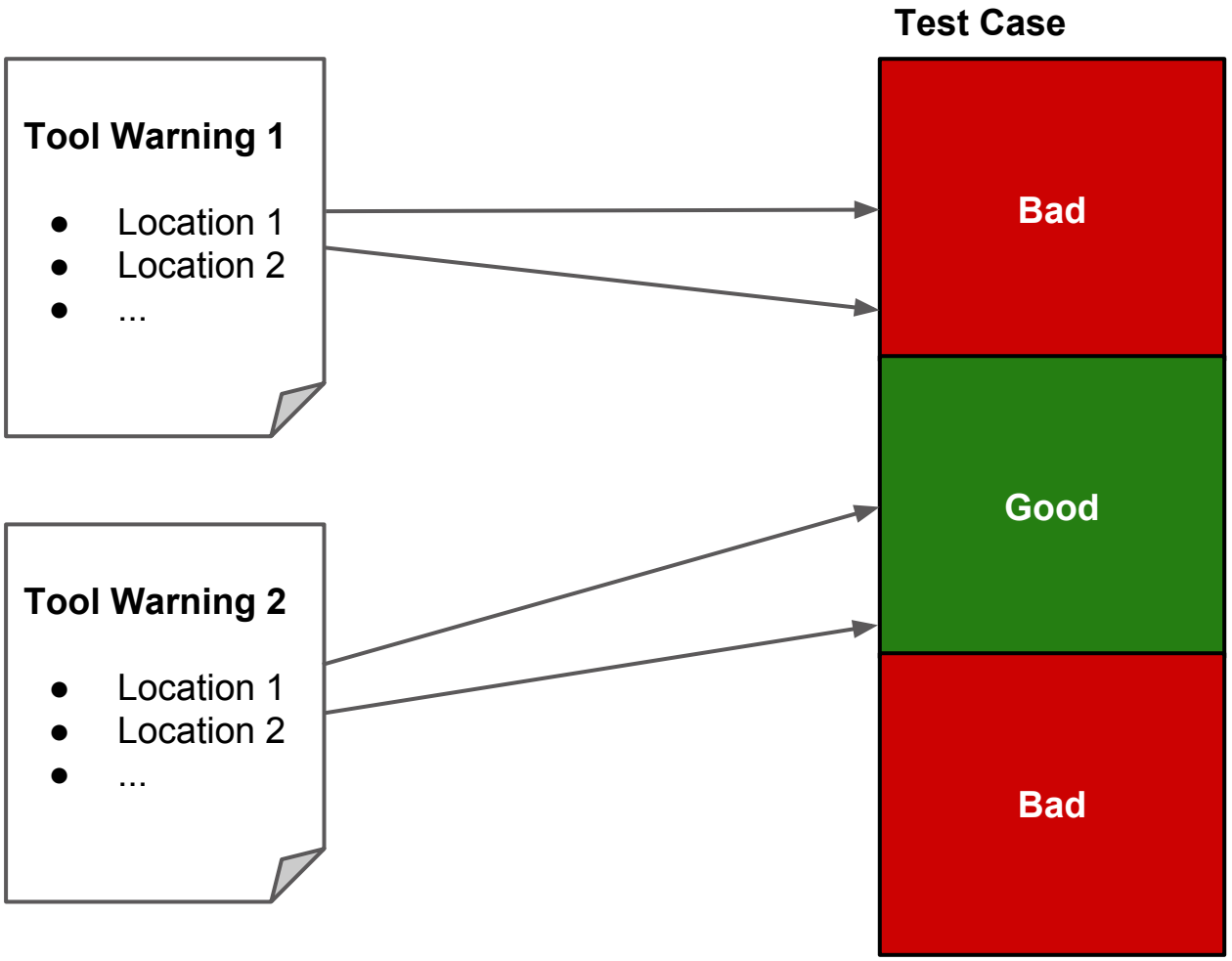
Tool Warning

- CWE List
- Path to Sink

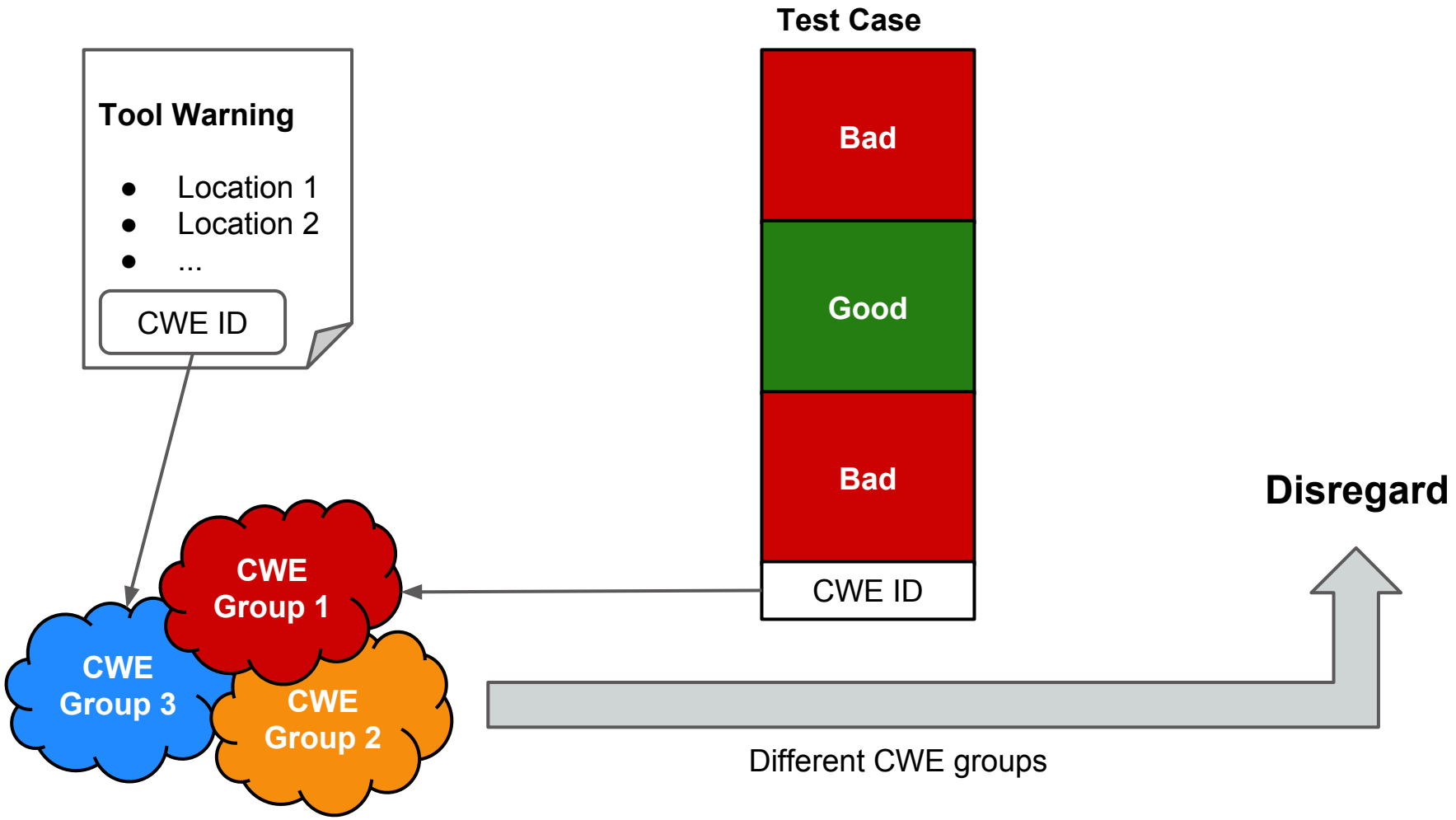
Test Case

- CWE
- Code Blocks
- Sinks

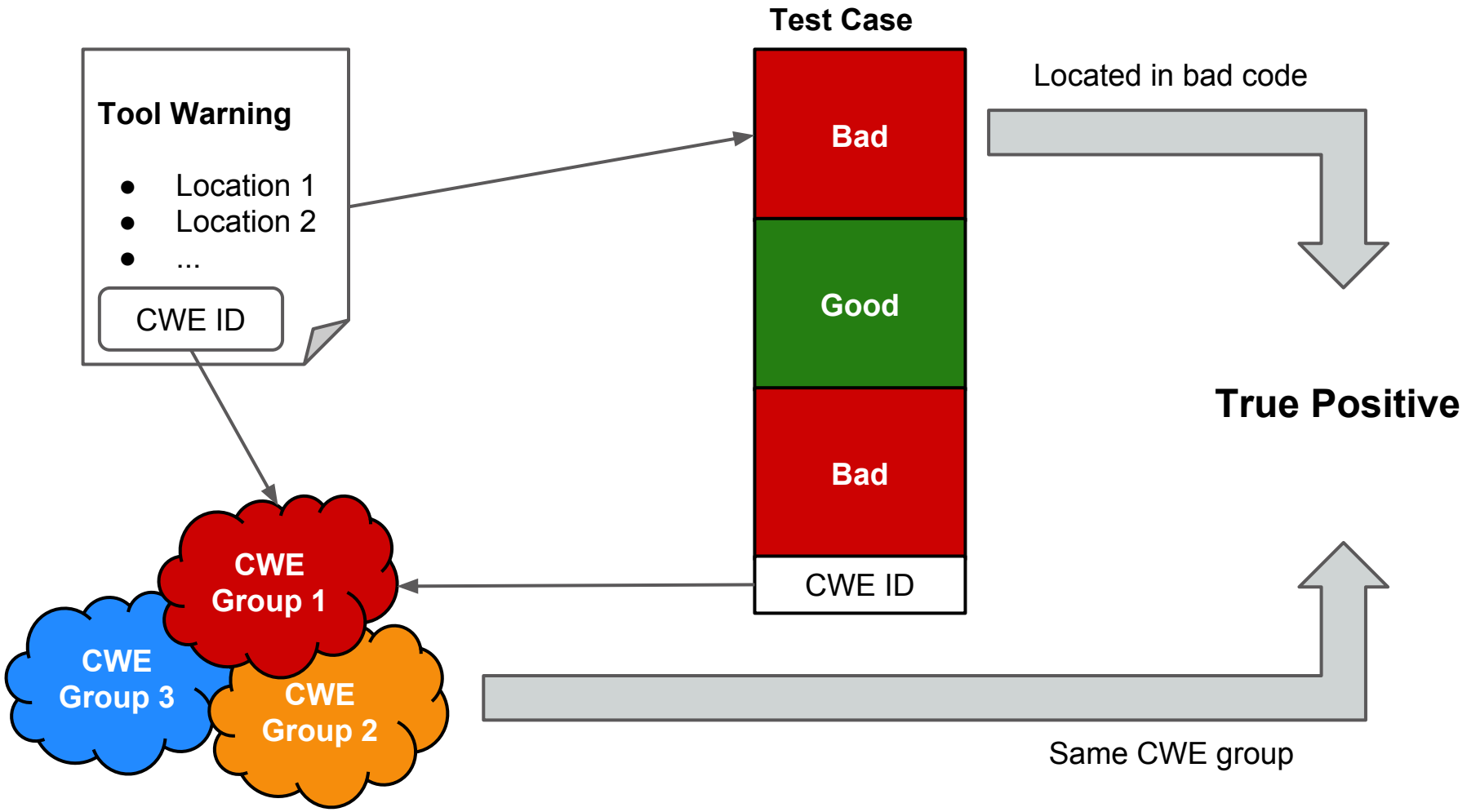
Location Matching



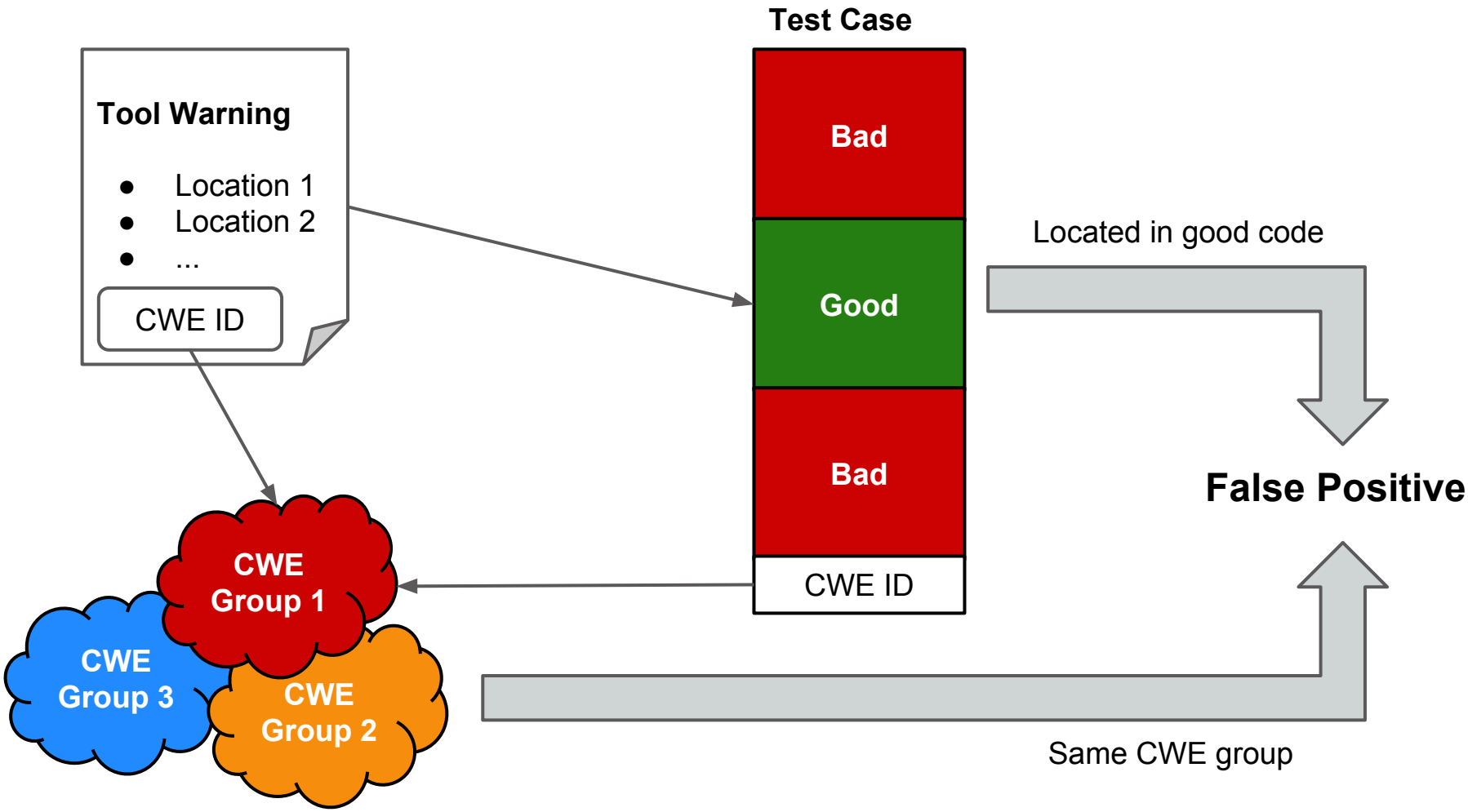
CWE + Location Supercombo



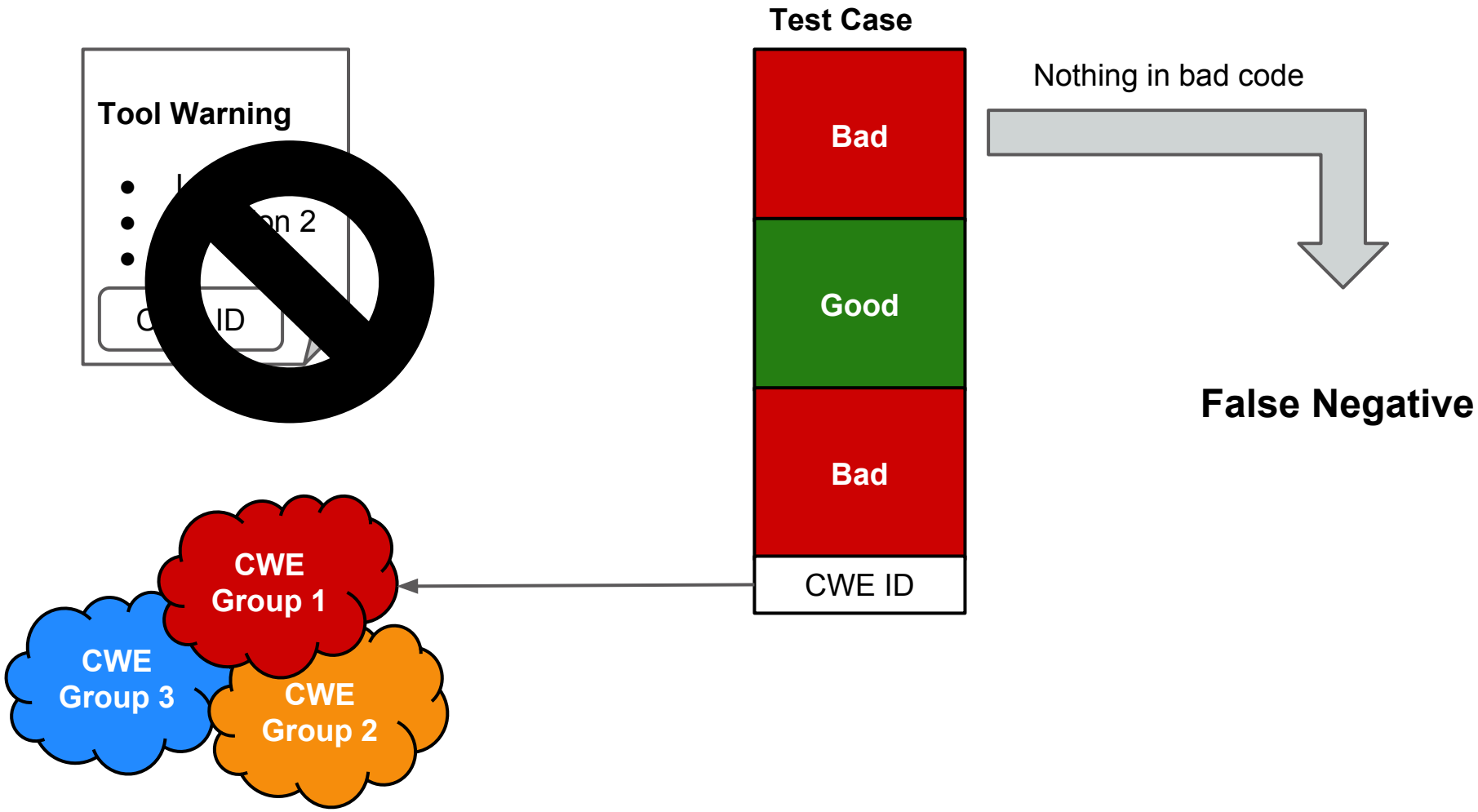
CWE + Location Supercombo



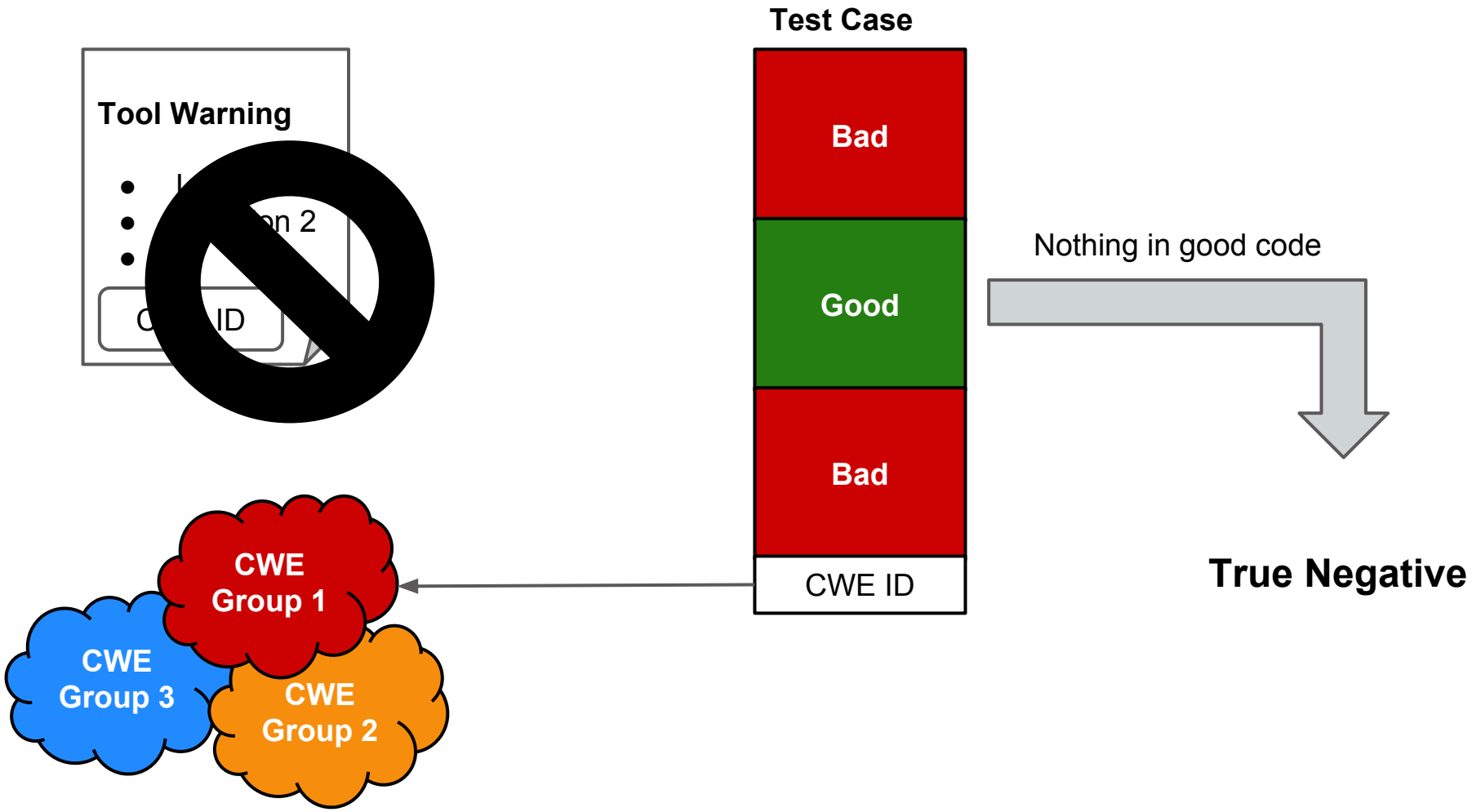
CWE + Location Supercombo



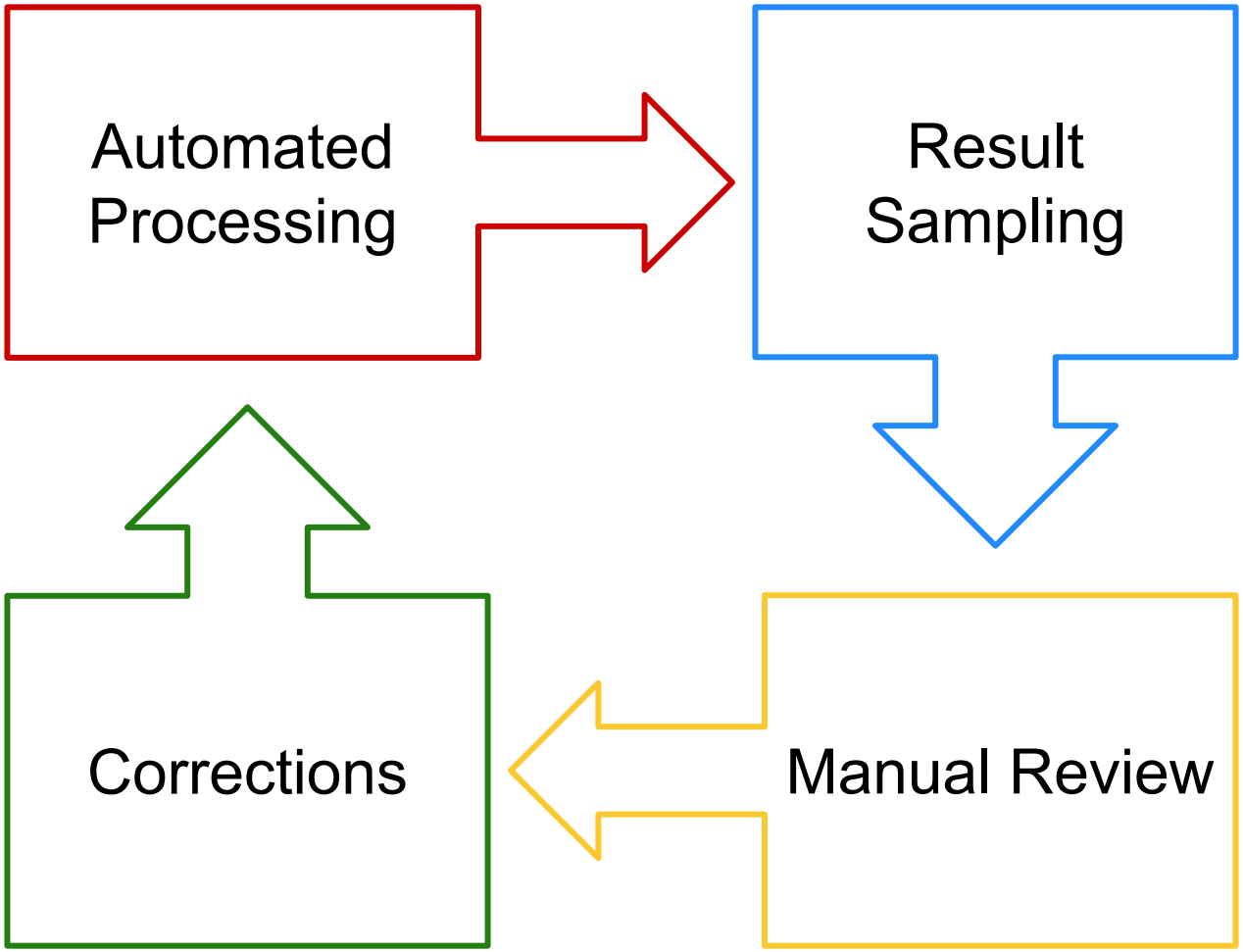
CWE + Location Supercombo



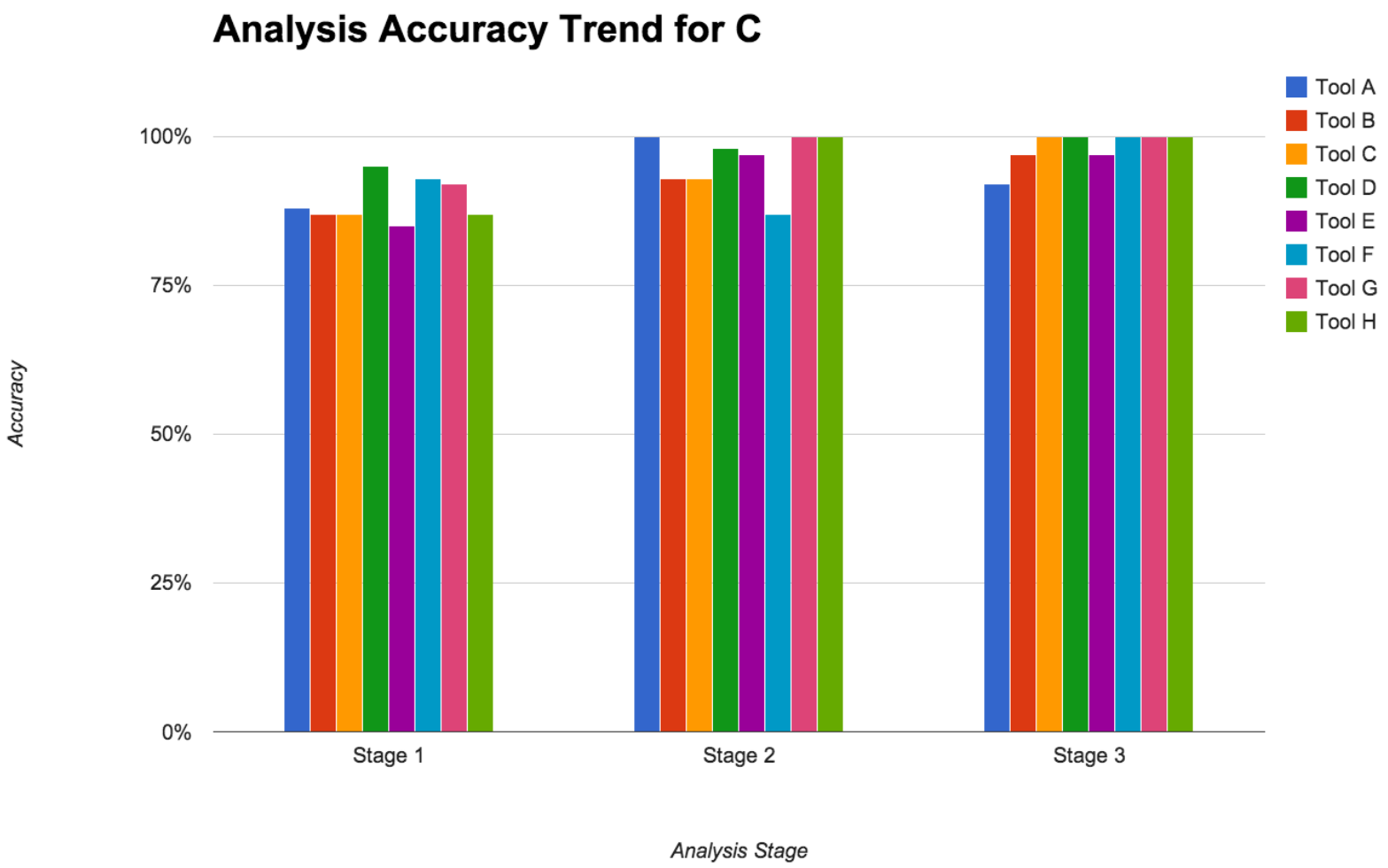
CWE + Location Supercombo



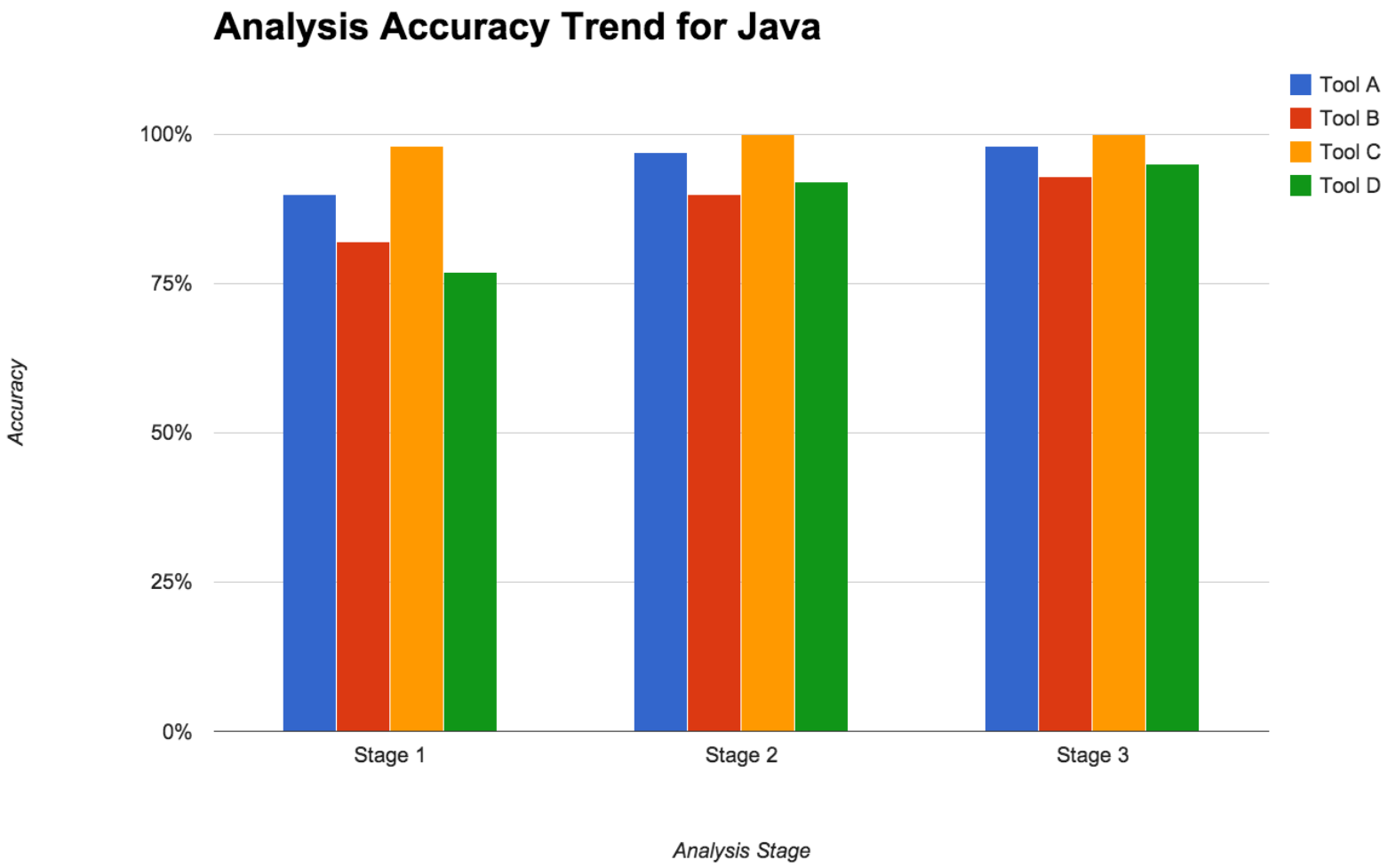
Analysis Cycle



Analysis Quality Trend (C)



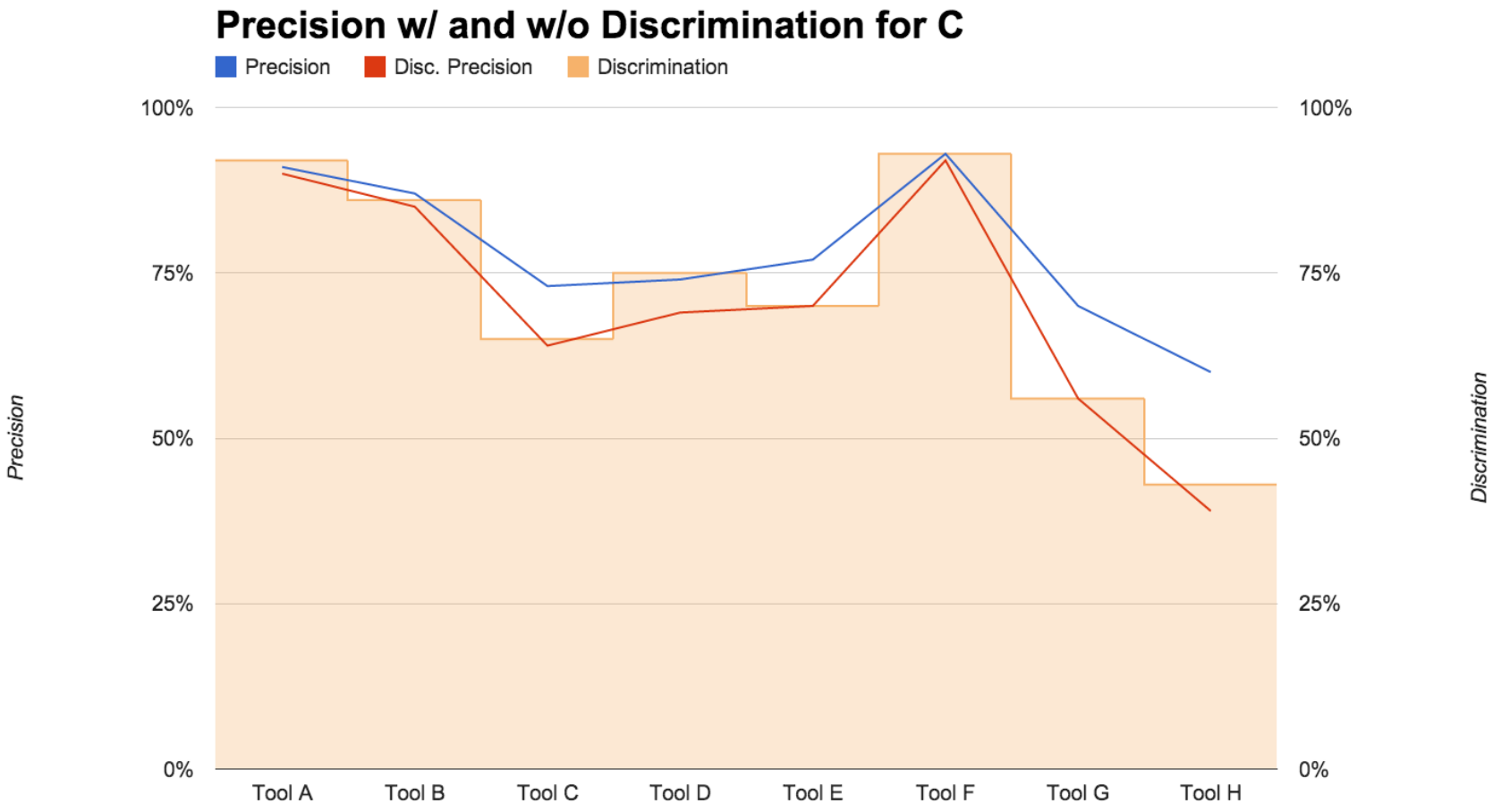
Analysis Quality Trend (J)



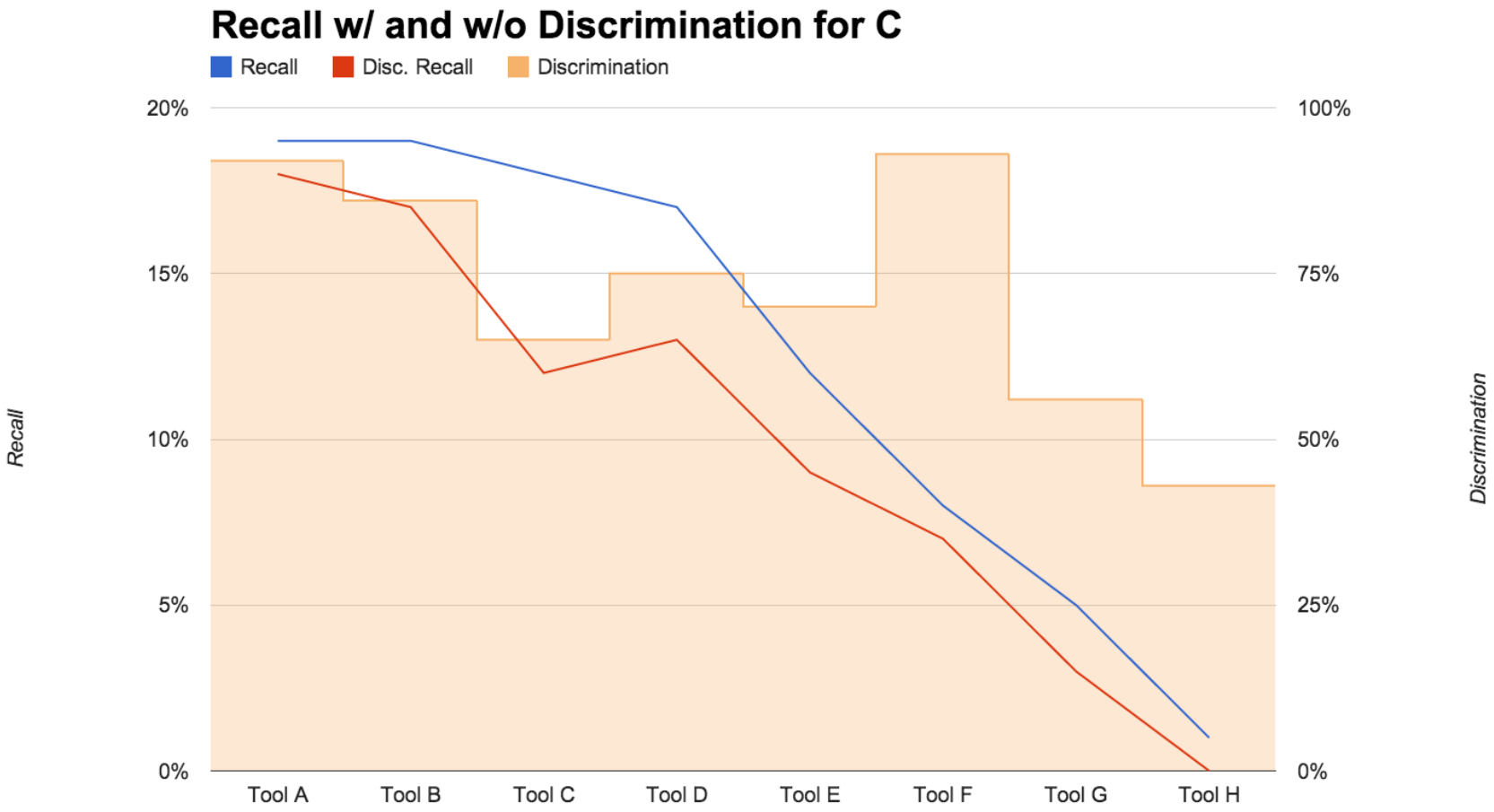
Metrics

Precision	How much can I trust a tool?
Recall	What proportion of flaws can a tool find?
Coverage	What kind of flaws can a tool find?
Discrimination	How smart is a tool?

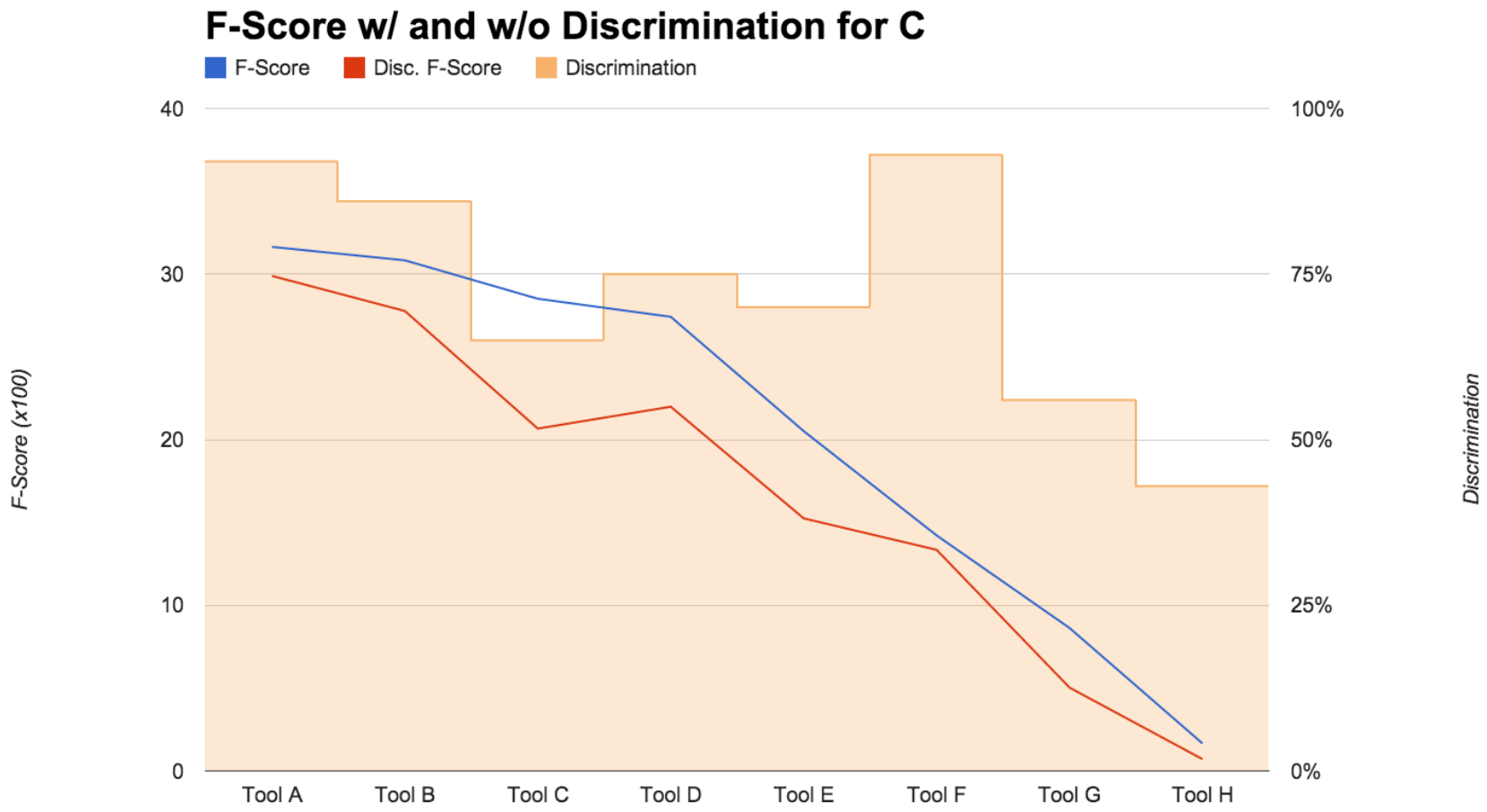
How much can I trust a tool? (C)



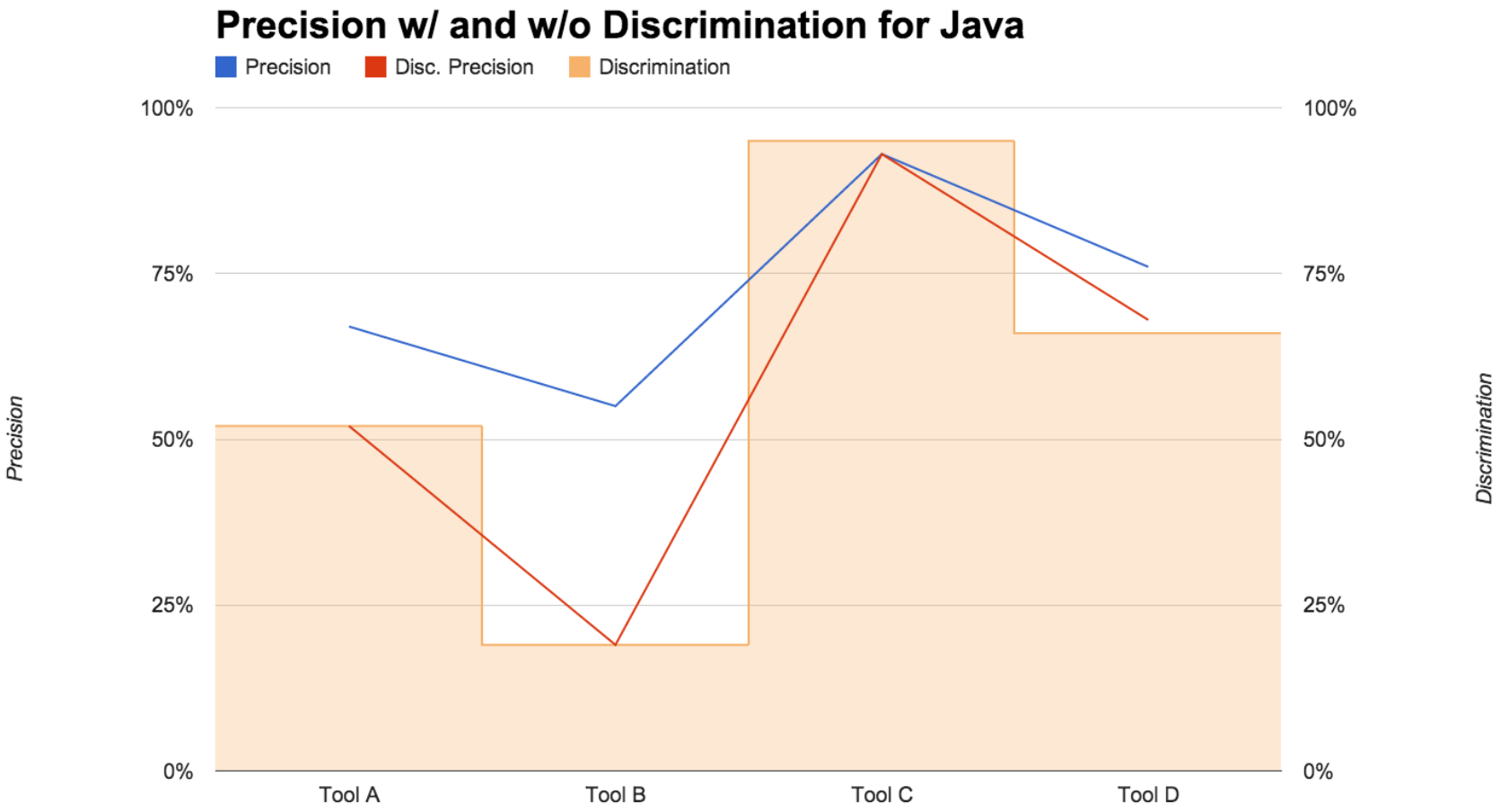
Proportion of flaws found by tools? (C)



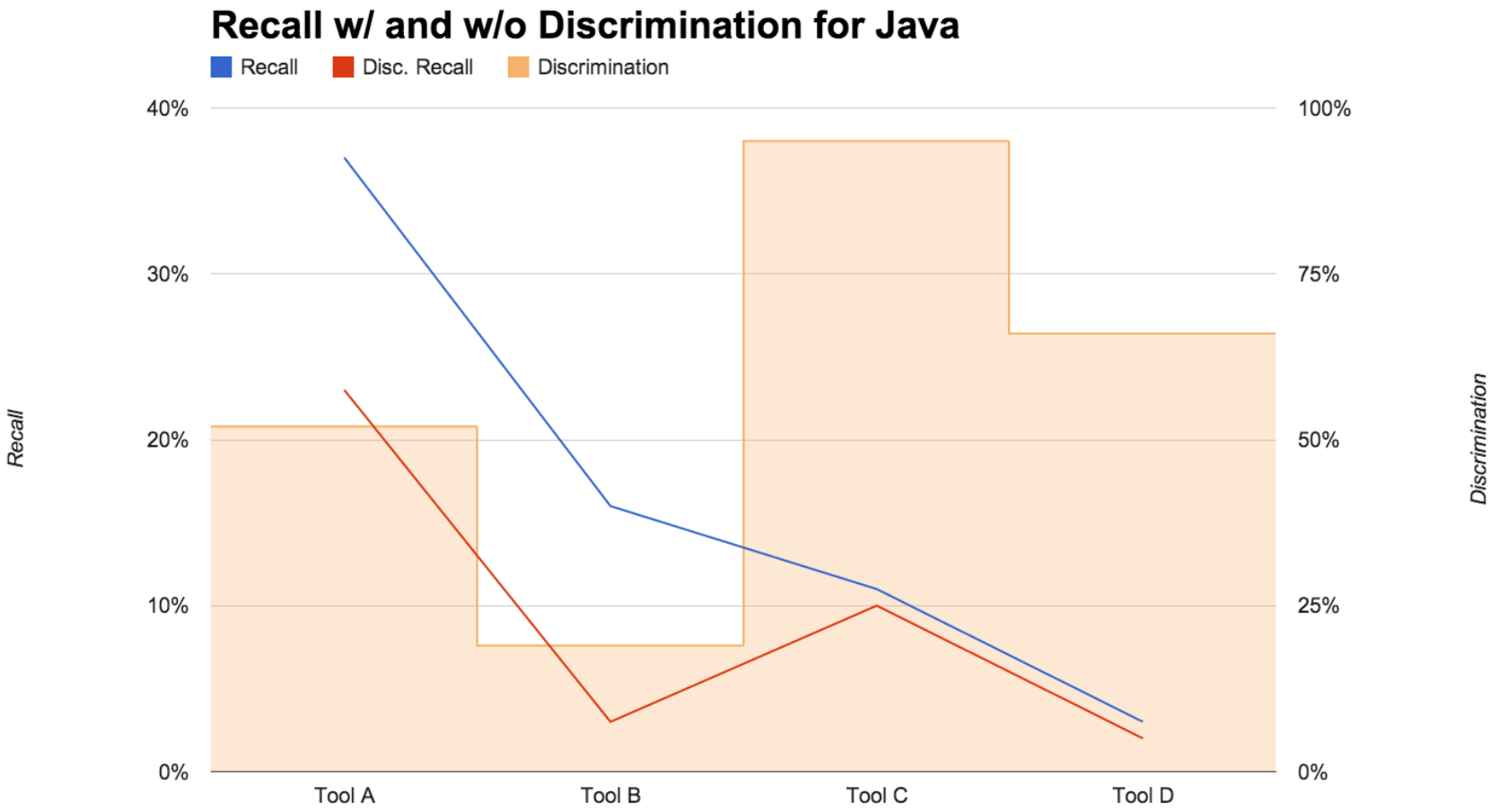
Tentative Overall Performance (C)



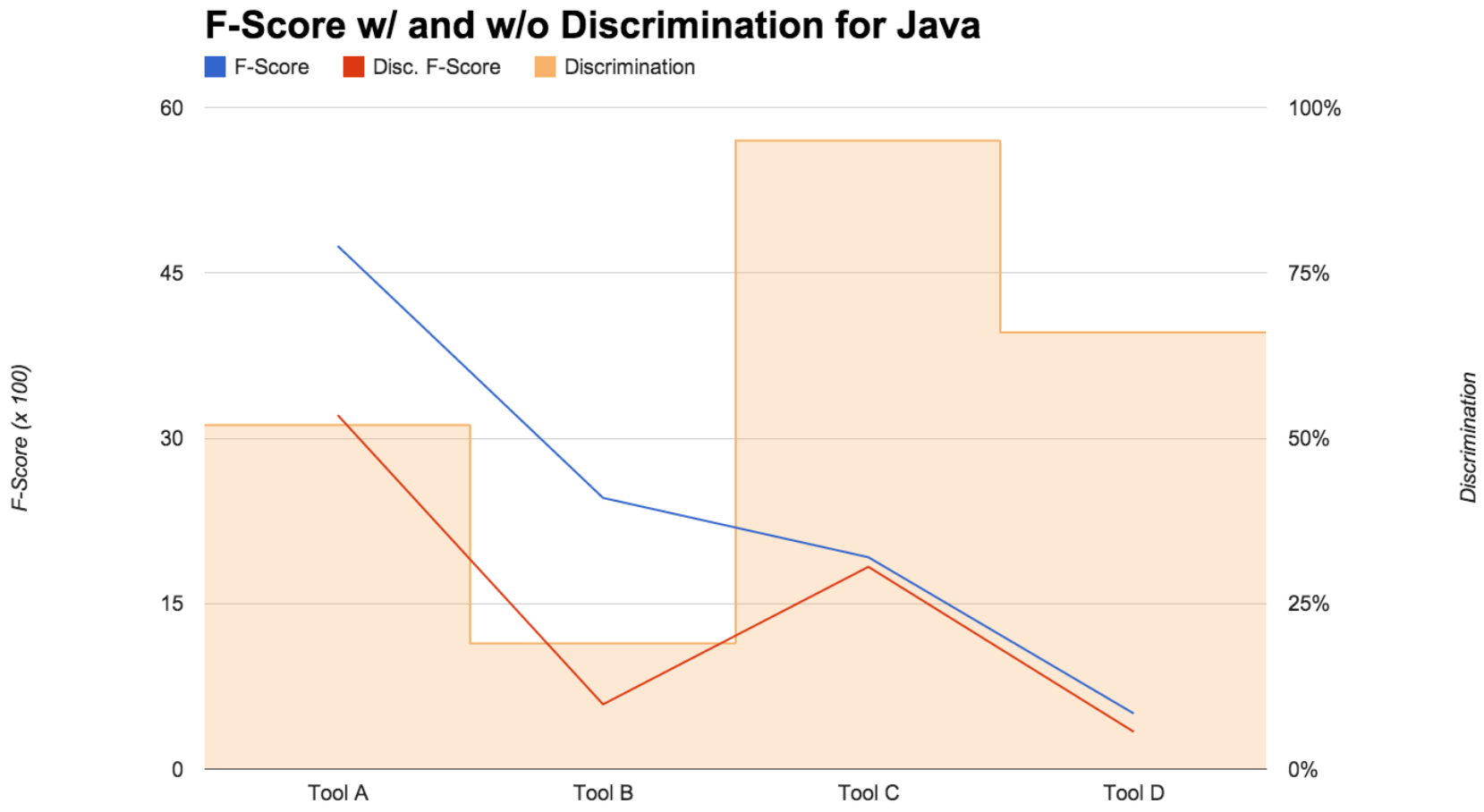
How much can I trust a tool? (J)



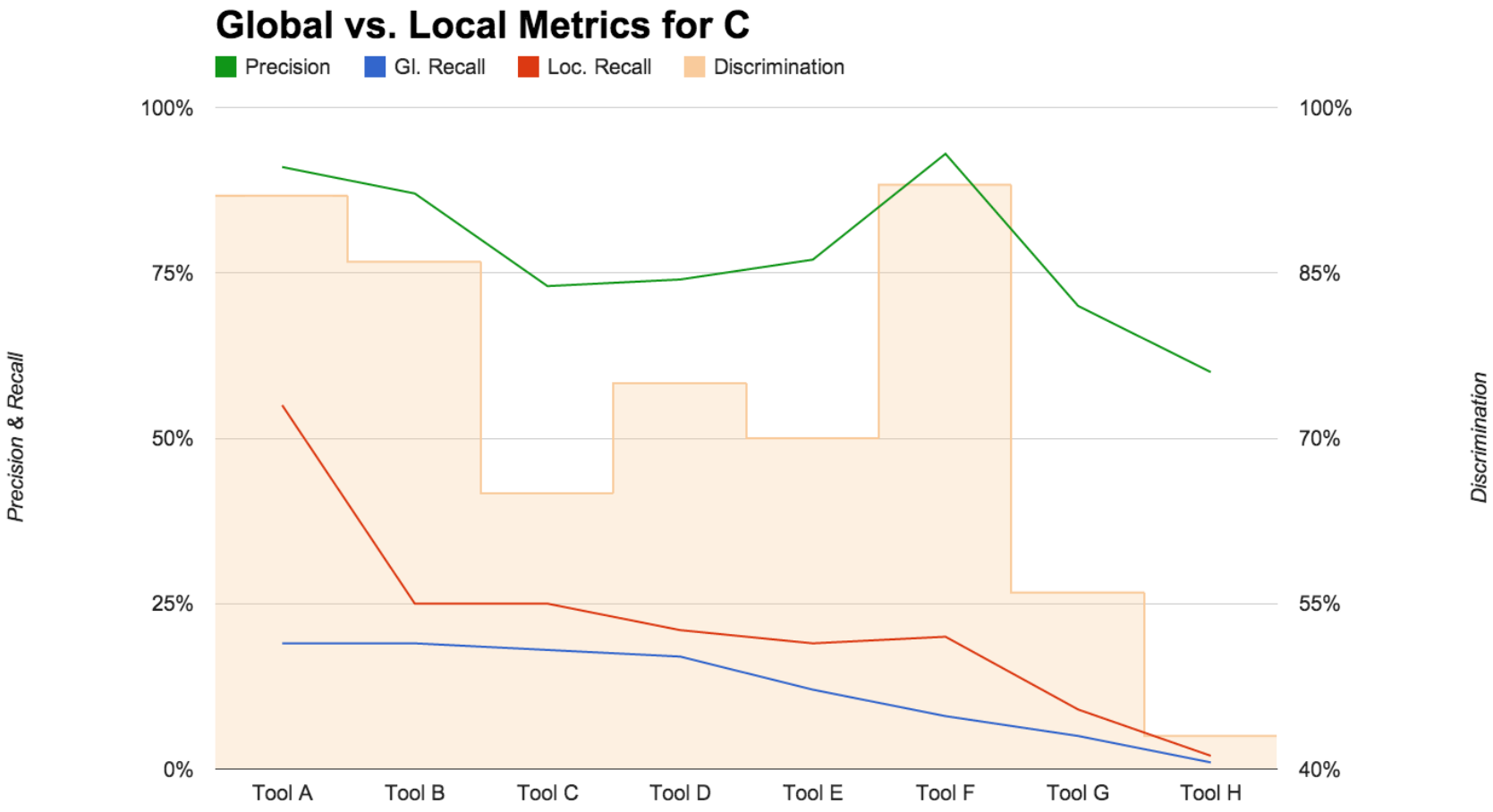
Proportion of flaws found by tools? (J)



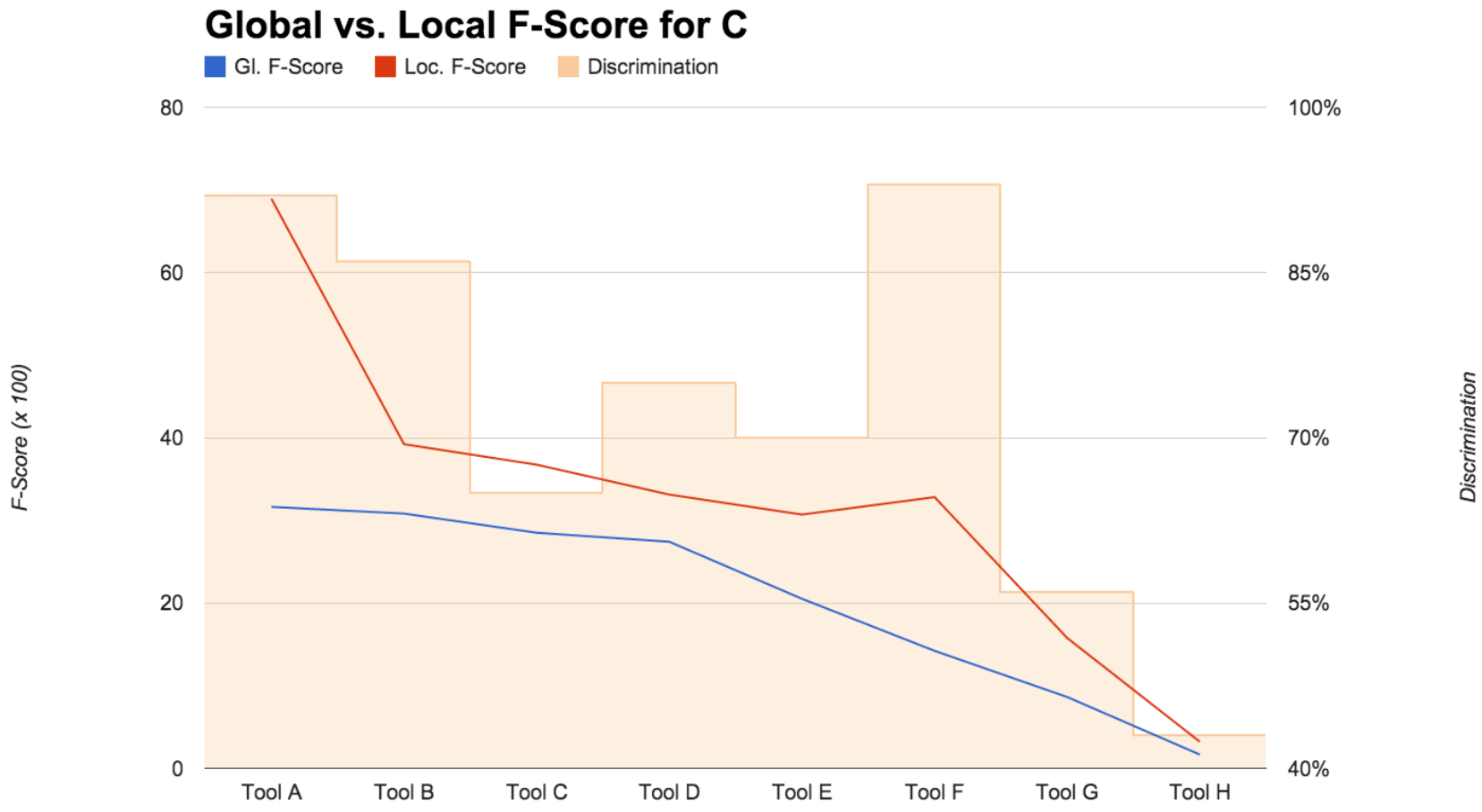
Tentative Overall Performance (J)



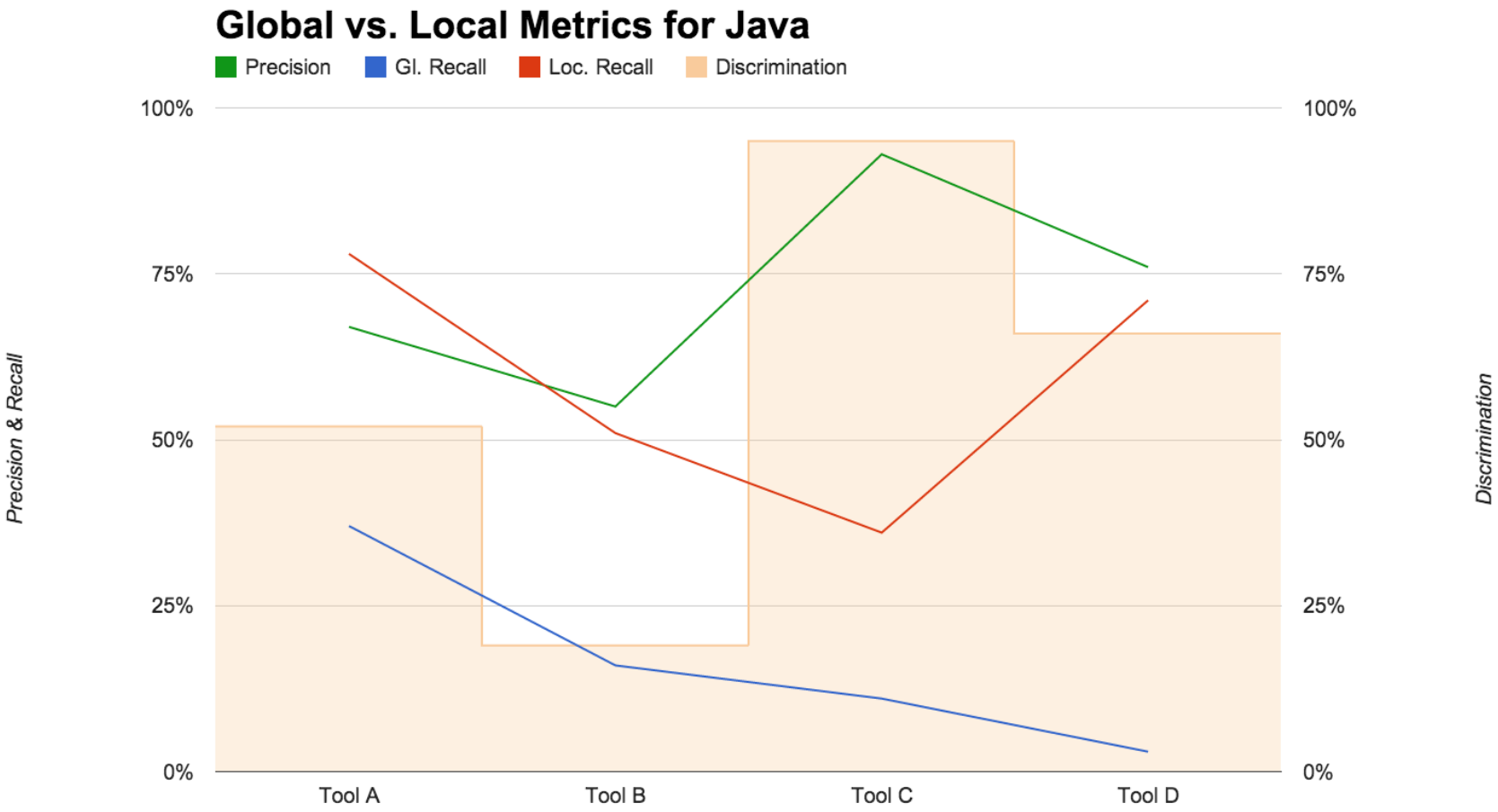
Global vs. Condensed (C)



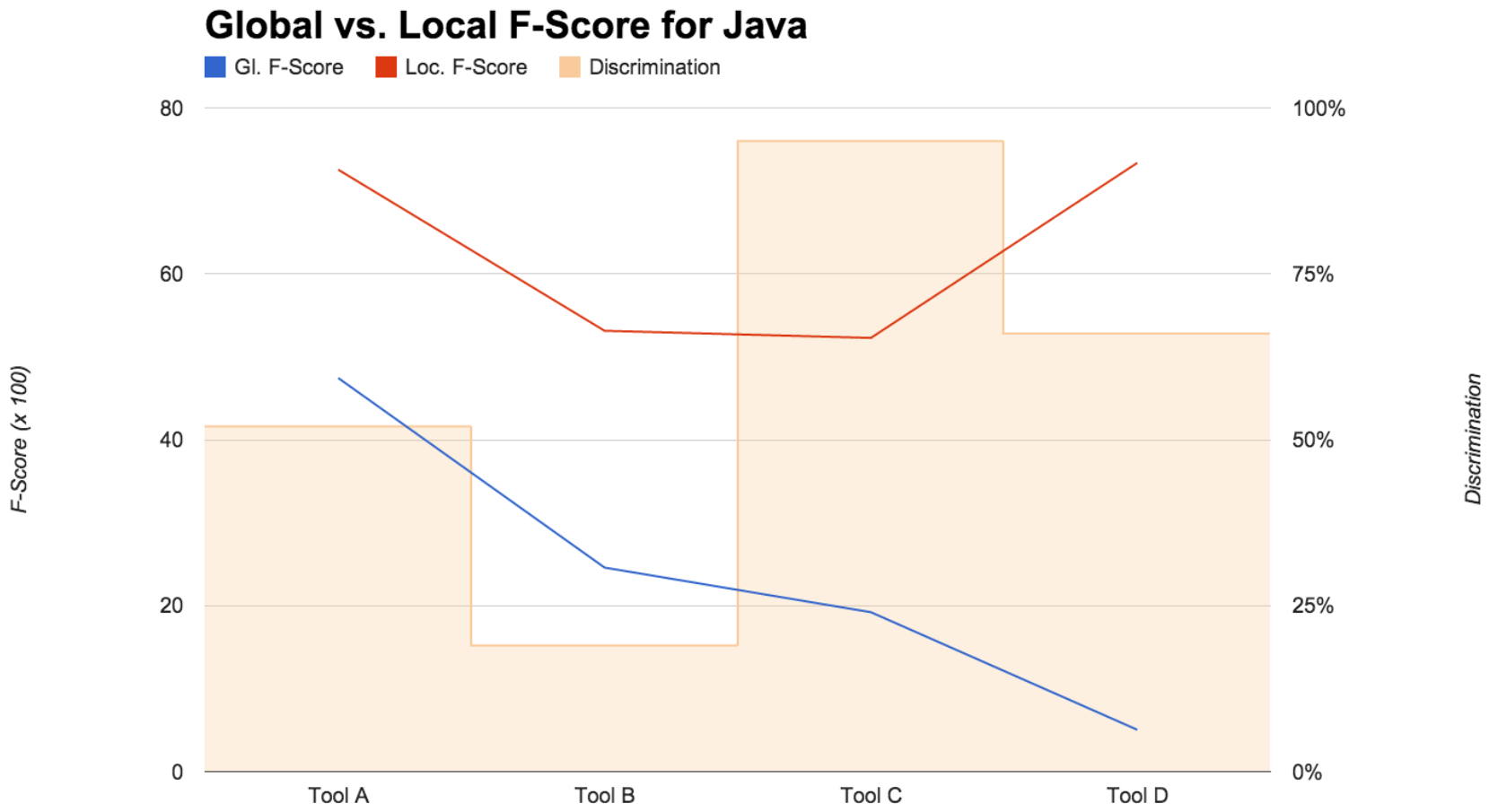
Global vs. Condensed (C)



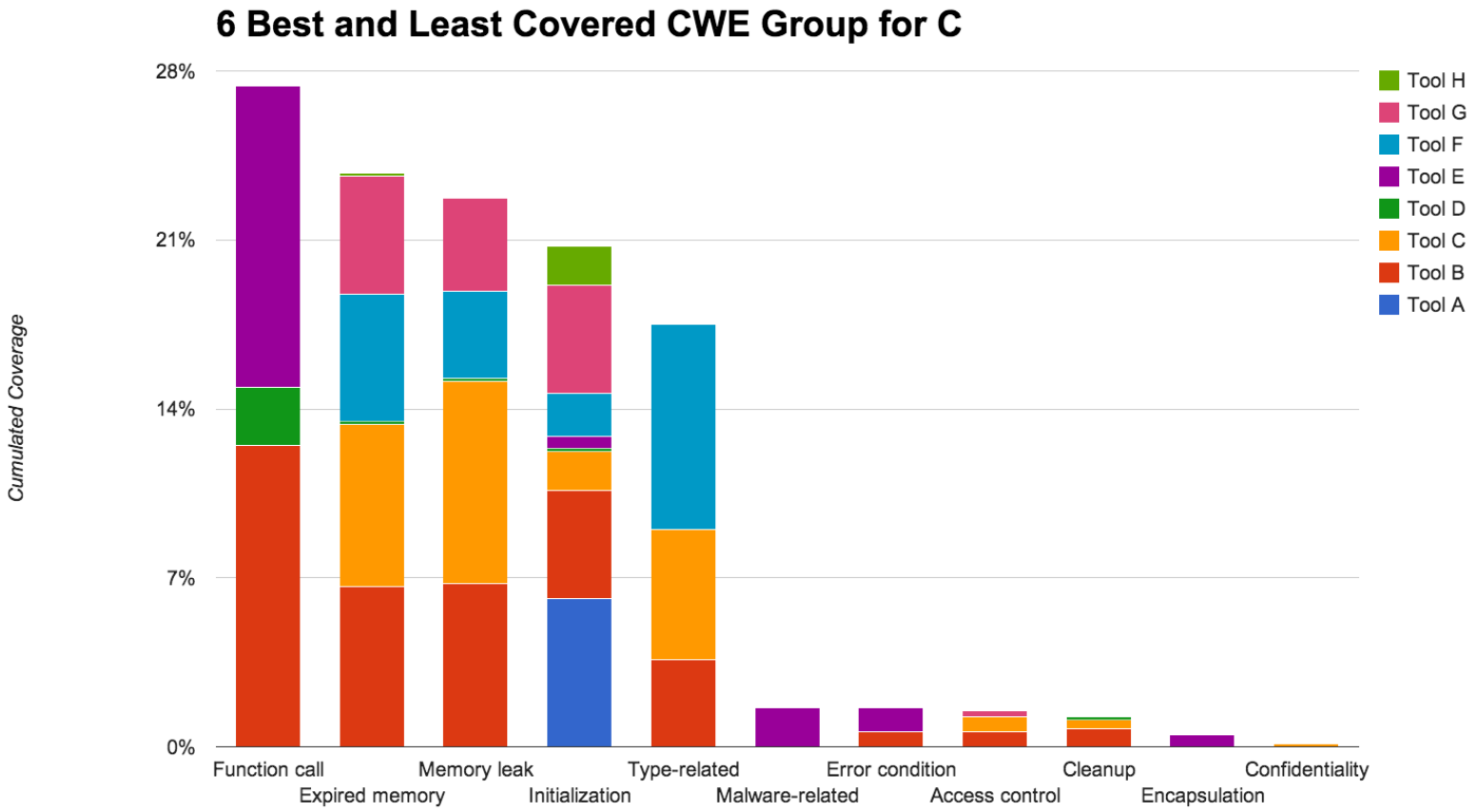
Global vs. Condensed (J)



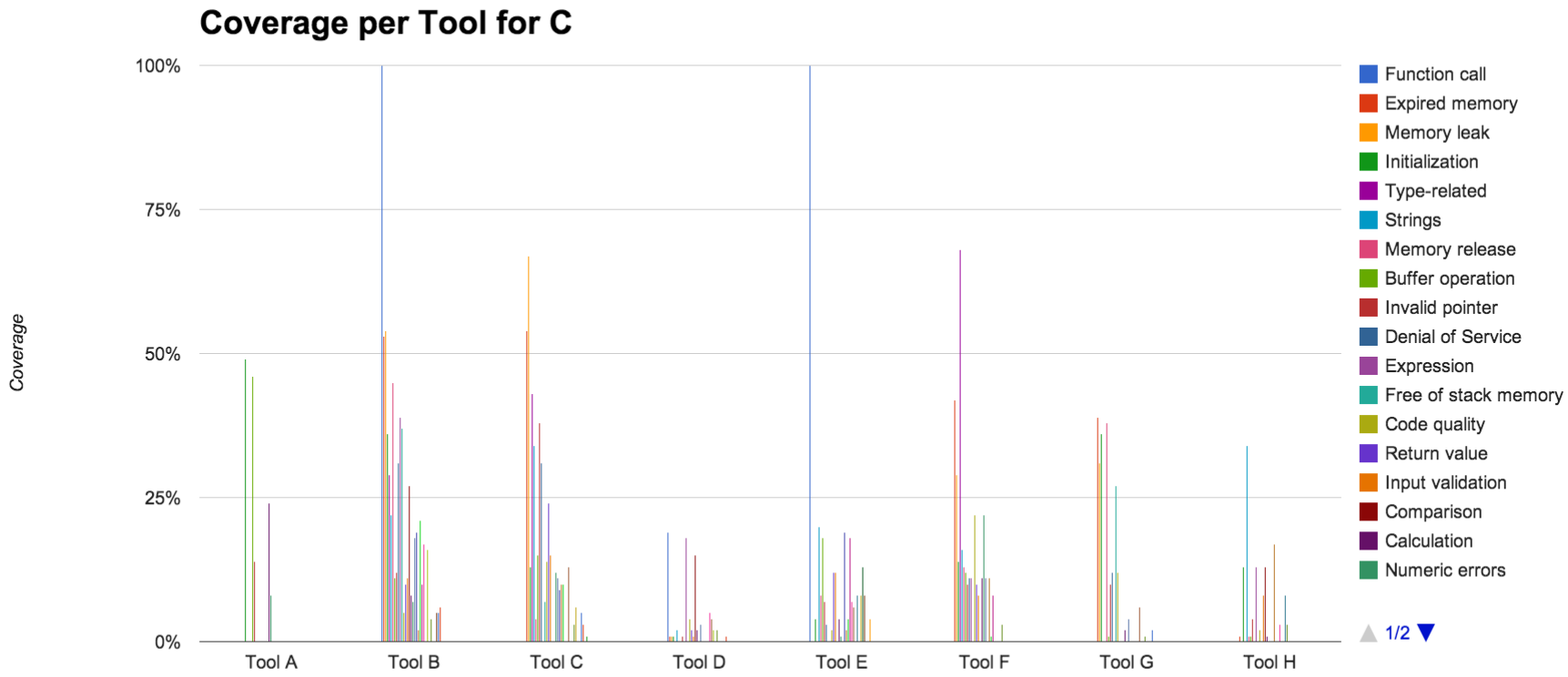
Global vs. Condensed (J)



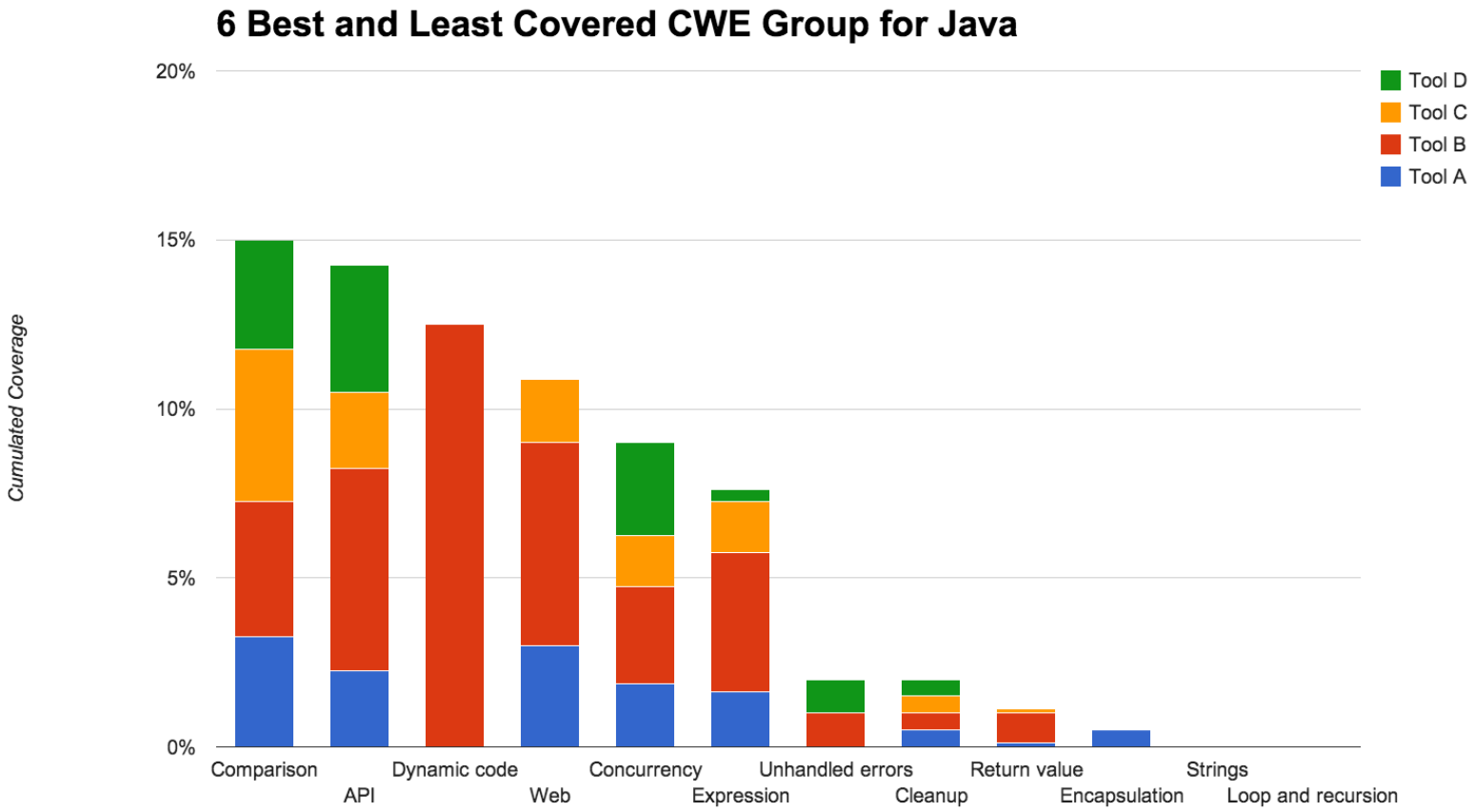
Best Covered Weakness Classes (C)



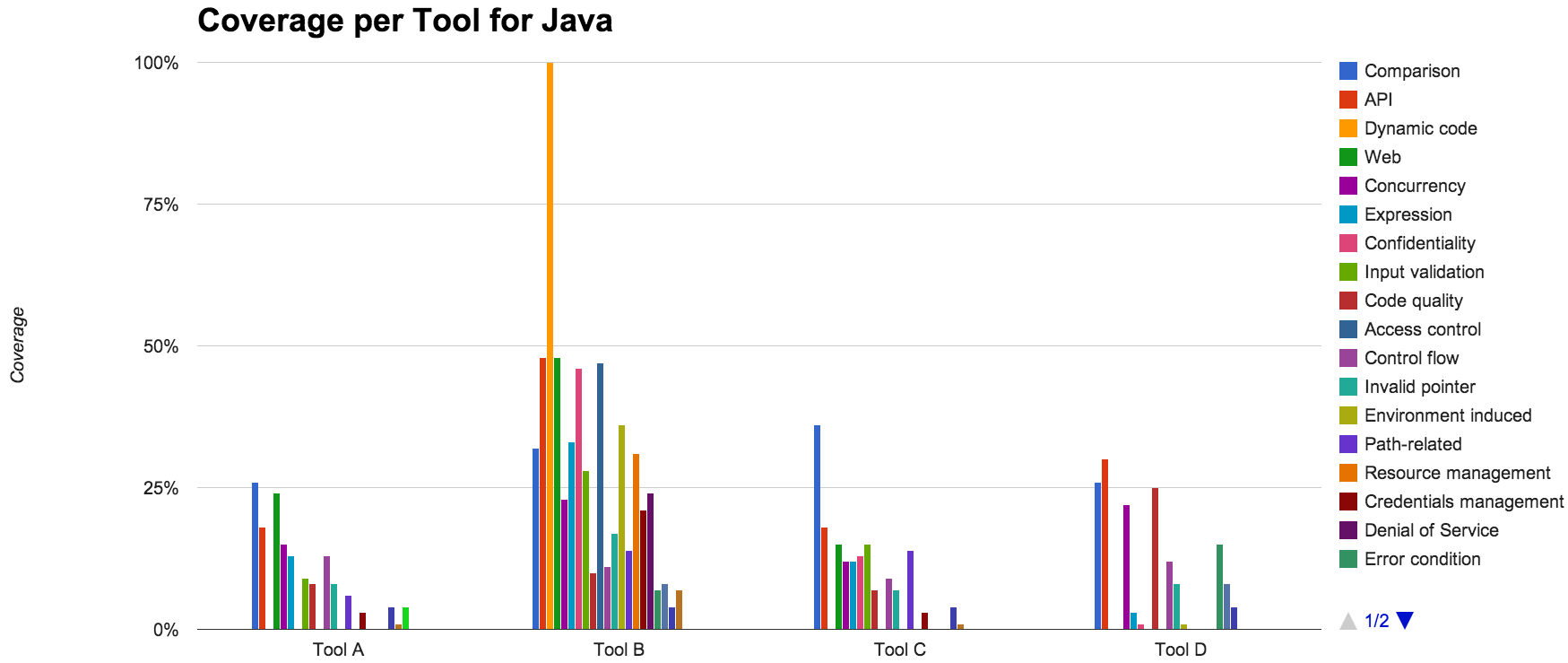
Coverage Spectrum (C)



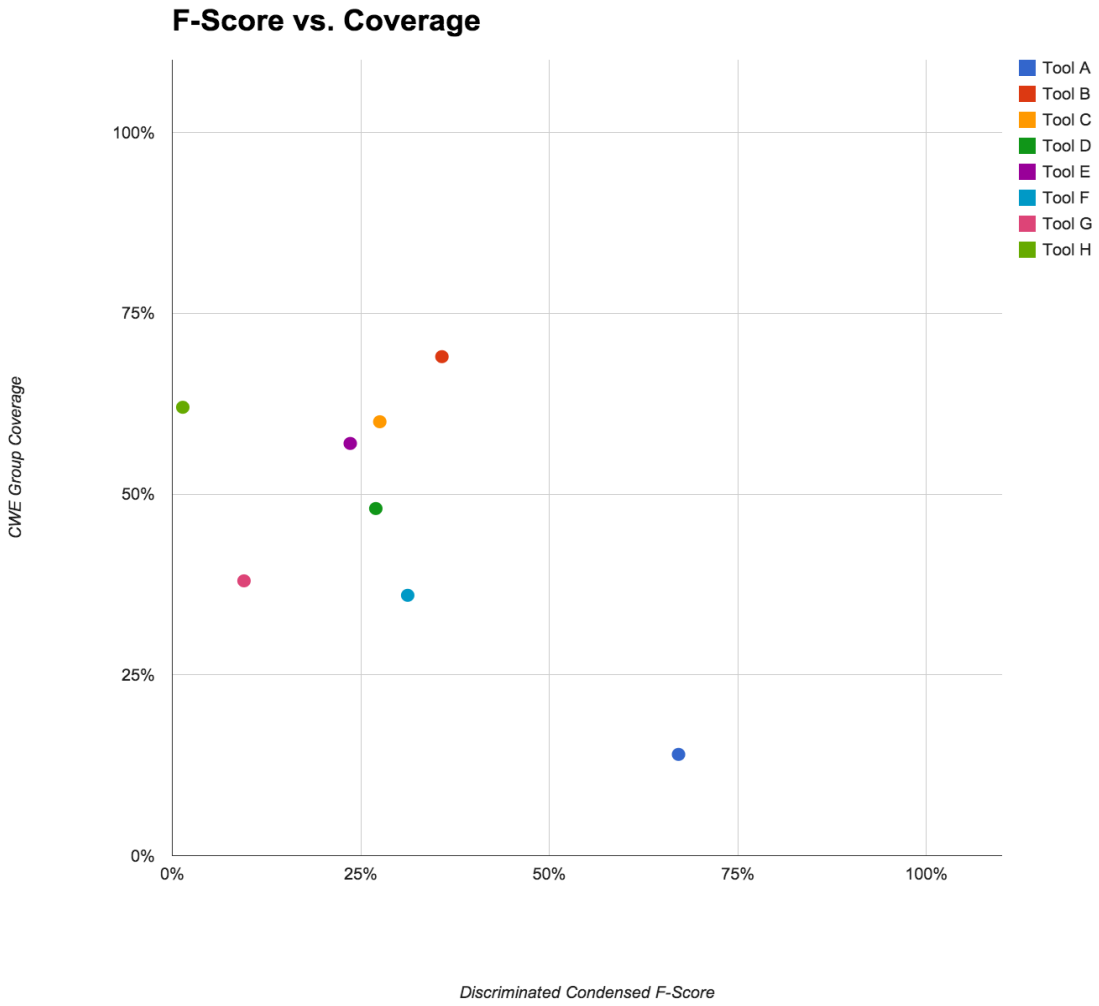
Best Covered Weakness Classes (J)



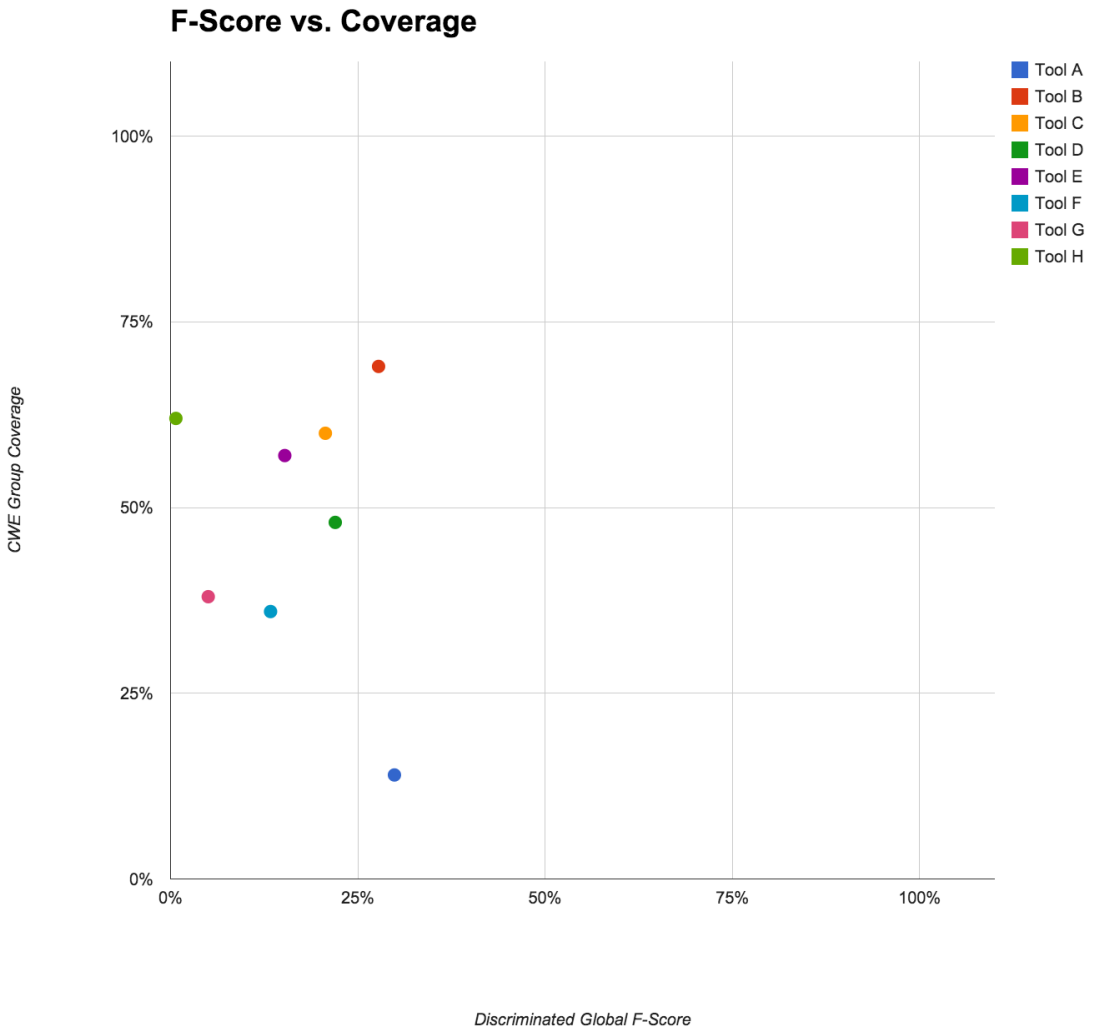
Coverage Spectrum (J)



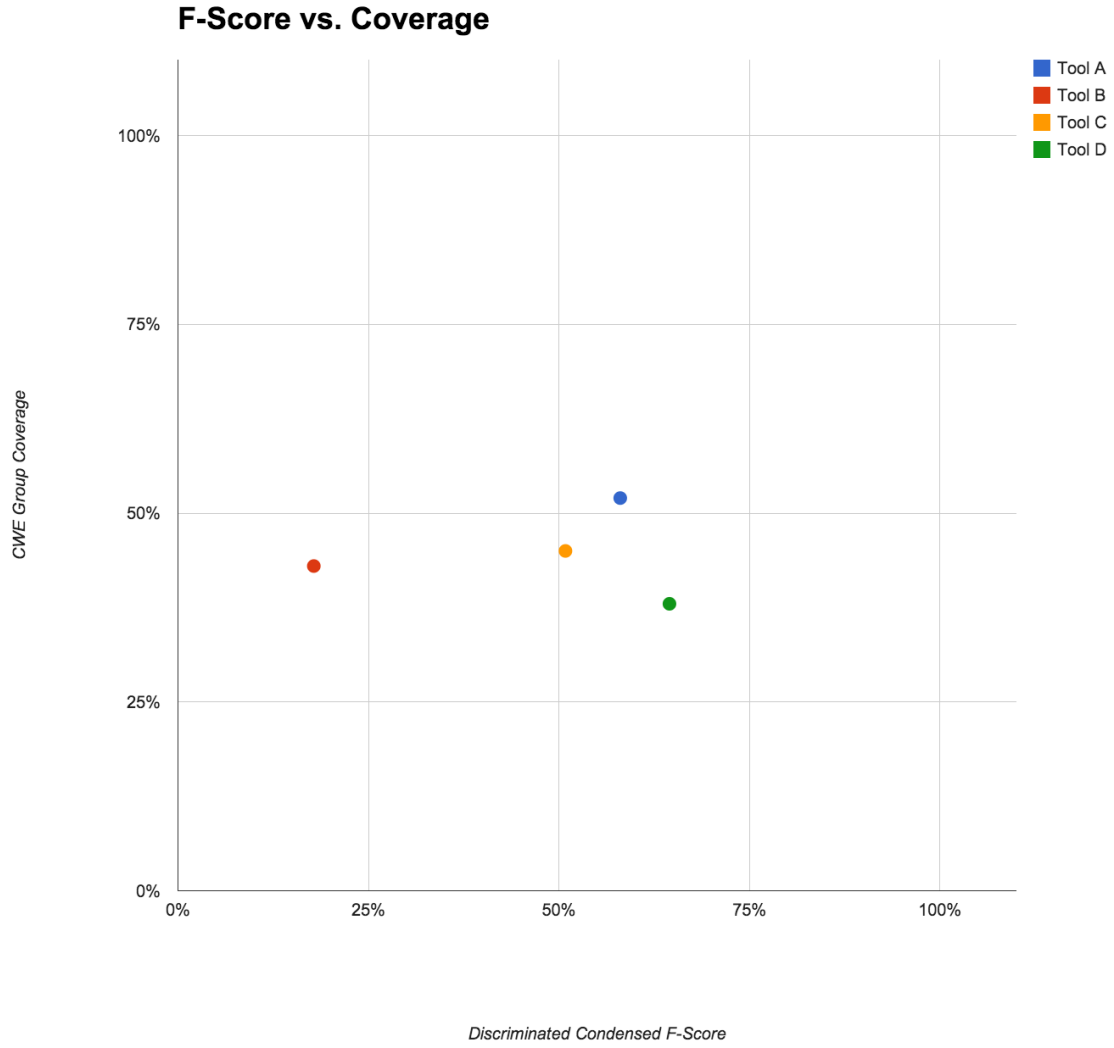
Overall Performance v2 (C)



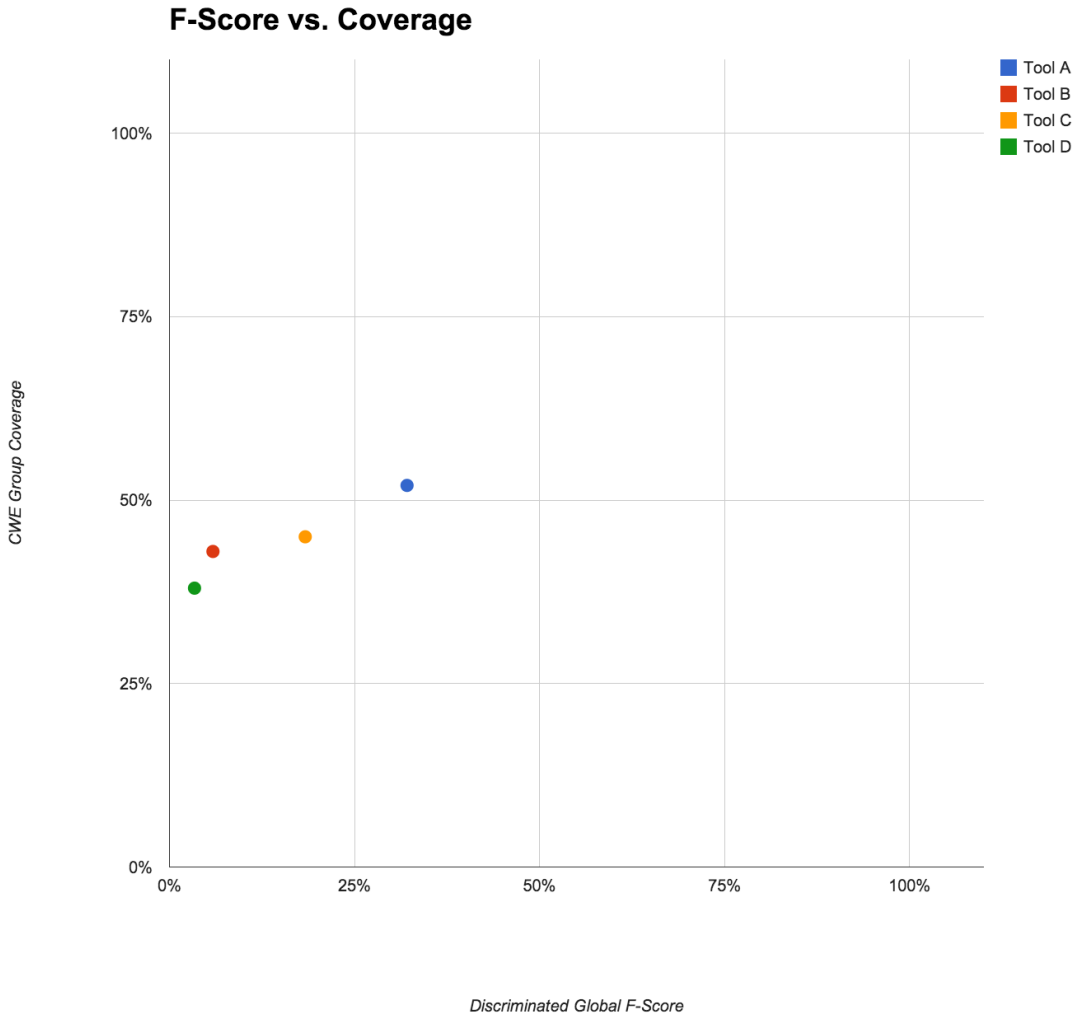
Overall Performance v2 (C)



Overall Performance v2 (J)



Overall Performance v2 (J)



Future Work

Metrics improvement

Introduce other aspects

- Test case complexity
- Overlap

More Cycles!

Conclusion

Tools differ in several dimensions

Metrics require careful development