

Deborah Mayo

Virginia Tech

The Statistics Wars and Their Casualties

High-profile failures of replication in the social and biological sciences underwrite a minimal requirement of evidence: If little or nothing has been done to rule out flaws in inferring a claim, then it has not passed a *severe test*. A claim is severely tested to the extent it has been subjected to and passes a test that probably would have found flaws, were they present. This probability is the severity with which a claim has passed. The goal of highly well-tested claims differs from that of highly probable ones, explaining why experts so often disagree about statistical reforms. Even where today's statistical test critics see themselves as merely objecting to misuses and misinterpretations, the reforms they recommend often grow out of presuppositions about the role of probability in inductive-statistical inference. Paradoxically, I will argue, some of the reforms intended to replace or improve on statistical significance tests enable rather than reveal illicit inferences due to cherry-picking, multiple testing, and data-dredging. Some preclude testing and falsifying claims altogether. These are the "casualties" on which I will focus. I will consider Fisherian vs Neyman-Pearson tests, Bayes factors, Bayesian posteriors, likelihoodist assessments, and the "screening model" of tests (a quasi-Bayesian-frequentist assessment). Whether one accepts this philosophy of evidence, I argue, that it provides a standpoint for avoiding both the fallacies of statistical testing and the casualties of today's statistics wars.