This paper presents the pitfalls of statistically significance testing and its effects on predicting causal inference, especially in toxicology and biomedical sciences. Statistical testings are often used to make inferences about causal relations to inform the risk assessment of interventions such as drugs or of environmental chemicals.

Statistically significant testing mainly involves formulation of a null hypothesis which denotes that there is no significant causality between a chemical and its exposure effects based upon the determination of a p-value. If the p-value is less than 0.05 or in some cases 0.001, the null hypothesis is discarded. In other words, only when the p-value is less than the threshold, we can draw causal relationship between, for instance, a toxic chemical and its effect on health.

The main criticism against statistically significant testing is that the threshold values used in the testing is arbitrarily set, and that it is a ‘black and white’ testing. Only the comparison against the threshold is taken into account and the actual p-value is disregarded.

While many statisticians and official agencies have warned against the pitfalls, it took almost a century to officially “highlight these flaws and the most serious implications exerted, for instance, in toxicological risk assessment and in the establishment of causality in legal evaluations”. It was only in 2016 that American Statistical Association made an official statement recognizing and highlighting the problem. In a recent seminal paper publish in Nature, the view against statistically significant testing received support of a large number of scientists from many disciplines all over the world. Recently, many Editors of scientific journals in the field of epidemiology and public health, medicine, and psychology have accordingly decided to *ban or to discourage the reporting of the results as related to “statistical significance testing”* (Filippini et al. (2022)). In this instance, the approach taken by the US Supreme Court in the *Matrixx v. Syracusan* case appears to be scientifically sound and even anticipated by several scientific communities. The demise of this simplistic approach appears to be fully justified in favor of a more rigorous and methodologically correct method based upon a comprehensive assessment of the strengths and limitations of the available edivence.

Reference

Filippini, T. and Vincenti S.R., (2022), The role of statistical significance testing in public law and health risk assessment, *J Prev Med Hyg*. 2022 Mar; 63(1): E161–E165.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9121665/>