Exploratory multiple testing

Aldo Solari

Ca' Foscari University of Venice

In confirmatory data analysis, scientific questions are specified in advance, translated into null hypotheses before data collection and then tested while controlling the probability of any false discovery. This stringent approach, known as familywise error rate (FWER) control (Tukey, 1953), is well suited for settings with a small number of carefully chosen hypotheses and well-designed studies, but it becomes overly conservative and impractical in modern large scale data analysis driven by rapid data growth and increased computational power.

As a result, multiple testing has increasingly shifted toward exploratory data analysis, where large collections of hypotheses are tested to generate new scientific insights. This transition —from hypothesis confirmation in small scale studies to hypothesis generation in large-scale settings— has motivated the development of new methods. A key breakthrough is the false discovery rate (FDR), defined as the expected proportion of false discoveries among all discoveries, which provides a more suitable error criterion, together with the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995) for controlling it.

Beyond controlling the FDR for a single set of discoveries produced by BH, recent work has focused on assessing the false discovery proportion (FDP) for sets of discoveries selected in a data-driven, post hoc manner. Goeman and Solari (2011) introduced a framework for post hoc FDP control based on the closure principle (Marcus, Peritz, and Gabriel, 1976), showing that closed testing underlies not only FWER control but also FDP control (Goeman, Hemerik, and Solari, 2021).

Until recently, however, no general framework combined FDR control with this flexible post hoc perspective. Recent work (Xu et al., 2025) addresses this gap by extending the closure principle using e-values, providing a general approach that allows flexible selection of discoveries and even post hoc choice of the error rate.

This talk will review these developments and present a unified perspective on multiple testing that bridges confirmatory and exploratory data analysis.

- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, Series B, 57, 289–300.

- Goeman, J. J., and Solari, A. (2011). Multiple testing for exploratory research. *Statistical Science*, 26(4), 584–597.

- Goeman, J. J., Hemerik, J., and Solari, A. (2021). Only closed testing procedures are admissible for controlling false discovery proportions. *The Annals of Statistics*, 49, 1218–1238.

- Marcus, R., Peritz, E., and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63, 655–660.

- Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished manuscript. In The Collected Works of John W. Tukey, Vol. VIII: Multiple Comparisons: 1948–1983 (pp. 1–300). Chapman and Hall, New York.

- Xu, Z., Solari, A., Fischer, L., de Heide, R., Ramdas, A., and Goeman, J. J. (2025). Bringing closure to false discovery rate control: A general principle for multiple testing. `arXiv:2509.02517`.