

## **Statistical boosting for biomedical research: strengths and limitations**

**Andreas Mayr (Universität Bonn, Alemania)**

Biostatisticians nowadays can choose from a huge toolbox of advanced methods and algorithms for prediction purposes. Some of these tools are based on concepts from machine learning; other methods rely on more classical statistical modelling approaches. In clinical settings, doctors are sometimes reluctant to consider risk scores that are constructed by black-box algorithms without clinically meaningful interpretation. Furthermore, even a both accurate and interpretable model will not often be used in practice, when it is based on variables that are difficult to obtain in clinical routine or when its calculation is too complex. In this talk, I will give a non-technical introduction to statistical boosting algorithms which can be interpreted as the methodological intersection between machine learning and statistical modelling. Boosting is able to perform variable selection while estimating statistical models from potentially high-dimensional data. It is mainly suitable for exploratory data analysis or prediction purposes. I will give an overview on some current methodological developments and provide an example for the construction of a clinical risk score. Another example will include the development of polygenic risk scores based on large genetic cohort data.