Meta-analysis and big data in the case of binary data

Stephan Morgenthaler*
Ecole polytechnique fédérale de Lausanne, Switzerland - stephan.morgenthaler@epfl.ch

I take the term big data to mean the analysis of data collected from disparate sources, some informative (but of varying quality), others heavily biased or even totally uninformative. One of the methods advocated for such data sets is to analyze smallish subsets of the data either randomly or systematically selected and then to combine the partial results into a global answer. This approach is superficially similar to a meta-analysis.
We will examine to what extent ideas and procedures from meta-analysis can be used in this context. The most important problem of meta-analyses is caused by the heterogeneity of the data, in particular various biases. The combination by meta-analysis across trials can produce biased answers associated with a false sense of accuracy. Similarly, in big data applications the combination of subset analyses can produce wrong answers.
We will take a straightforward example of a binary response variable and few covariates and/or factors as a basis for explaining the challenges and developing robust meta-analytic summaries.

Key Words: big data, binary regression, subset analyses, meta-analysis.