

September 2021

Paper of the Month, September 2021: Rubin, D. (1976). [Inference and missing data](#). *Biometrika*, 63(3), 581-592.

Incomplete data also refers to as missing data is a common complication in research. The impact of incomplete data is detrimental across most research. A simple Goggle Scholar search for missing data or incomplete data results in more than 5,000,000 hits. The basic theoretical structure for the analyses of incomplete data was envisioned and developed by Don Rubin in the early to mid 70's. Together with the EM Algorithm (Dempster, Laird & Rubin 1977) and Multiple imputation (Rubin 1977, Rubin 2004) the paper *Inference and missing data* has long lasting impact on research to this day.

We will discuss the basic ideas and implications of incomplete data together with concepts such as Missing at Random, Missing not at Random and ignorability which were coined by Rubin in this paper.

## References

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-22.

Rubin, D. B. (1977). The design of a general and flexible system for handling non-response in sample surveys," Consultant Report Submitted to the Social Security Administration, done as part of the 1973 CPS-IRS-SSA Exact Match Project. Also in *The American Statistician*, 58, 298–302.

Rubin, D. B. (2004). The design of a general and flexible system for handling nonresponse in sample surveys. *The American Statistician*, 58(4), 298-302.