Literature review

Hanne Oberman

17-8-2021

# Chapter text

Briefly, we searched relevant publications on PubMed and ArXiv that describe quantitative evaluations of missing data methods. Initially, we identified 15 relevant papers in which several statistical and machine learning imputation techniques were compared through simulation. These 15 papers served to inform a literature review using the active learning software ASReview [software ref]. ASReview is a machine-learning aided framework for screening (e.g.) titles and abstracts [nature ref].

* AI assisted reviewing, see Nature publication ASReview
* literature search PubMed and ArXiv with inclusion criteria: missing/incomplete, simulation
* manual selection of 15 relevant articles for stopping criterion AI assisted reviewing
* removed duplicates, start screening abstracts selected by ASReview
* stop screening when 15 manually selected articles are ‘found’ by ASReview
* screen full articles with inclusion criteria: missingness empirical/MAR/MNAR (not MCAR),
* flow chart of methods presented in Figure:

After omitting duplicates and removing papers that did not meet the inclusion criteria, we obtained 64 publications . These are listed on [<http://> xxxxx ].

## References

van de Schoot, R., de Bruin, J., Schram, R. et al. An open source machine learning framework for efficient and transparent systematic reviews. Nat Mach Intell 3, 125–133 (2021). www.nature.com/articles/s42256-020-00287-7

Van de Schoot, R., De Bruin, J., Schram, R., Zahedi, P., De Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Tummers, L., & Oberski, D. (2021). ASReview: Active learning for systematic reviews. Zenodo. <https://doi.org/10.5281/zenodo.5126631>

# Introduction

* missing data is problematic and inevitable when working with RWD in a clinical setting
* there has been a shift from CCA to stats to ML methods to handle missing data
* ML methods appear to be better suited for empirical data, where assumptions of CCA and stats methods do not hold
* this review aims to summarize simulation studies into missing data methods

# Methods

* AI assisted reviewing, see Nature publication ASReview
* literature search PubMed and ArXiv with inclusion criteria: missing/incomplete, simulation, …
* manual selection of 15 relevant articles for stopping criterion AI assisted reviewing
* removed duplicates, start screening abstracts selected by ASReview
* stop screening when 15 manually selected articles are ‘found’ by ASReview
* screen full articles with inclusion criteria: missingness empirical/MAR/MNAR (not MCAR), …
* flow chart of methods presented in Figure:

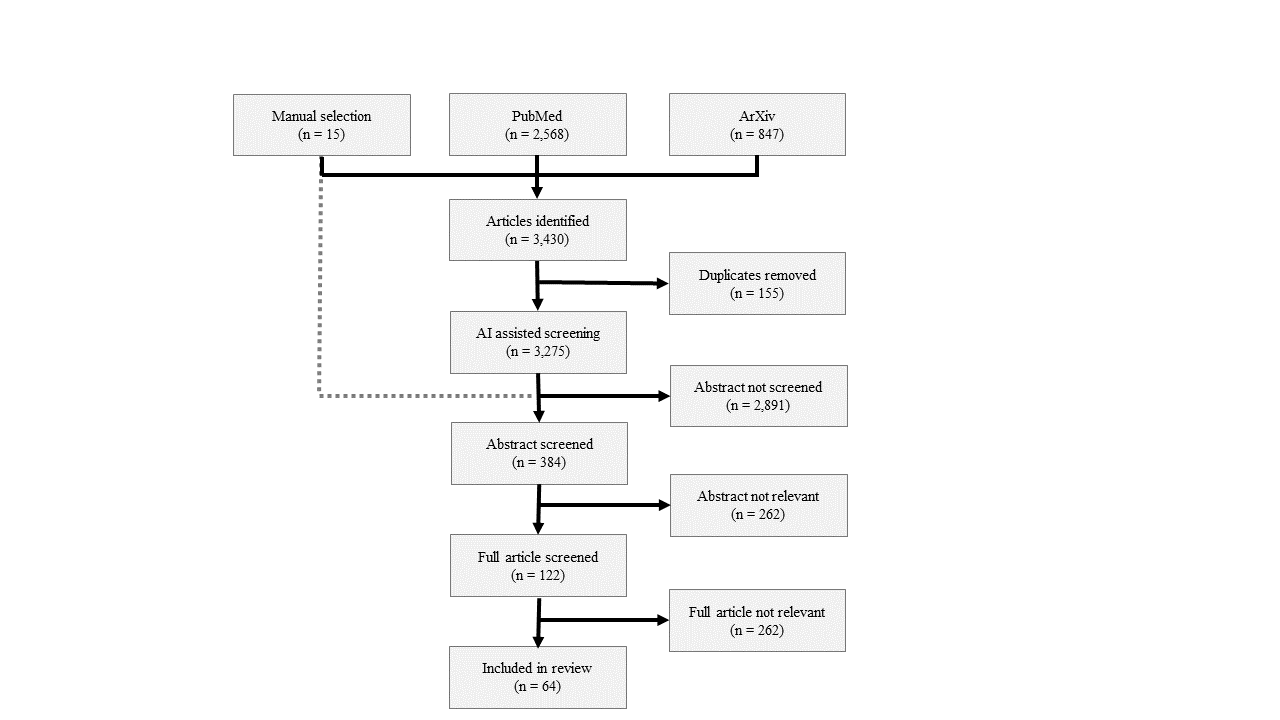
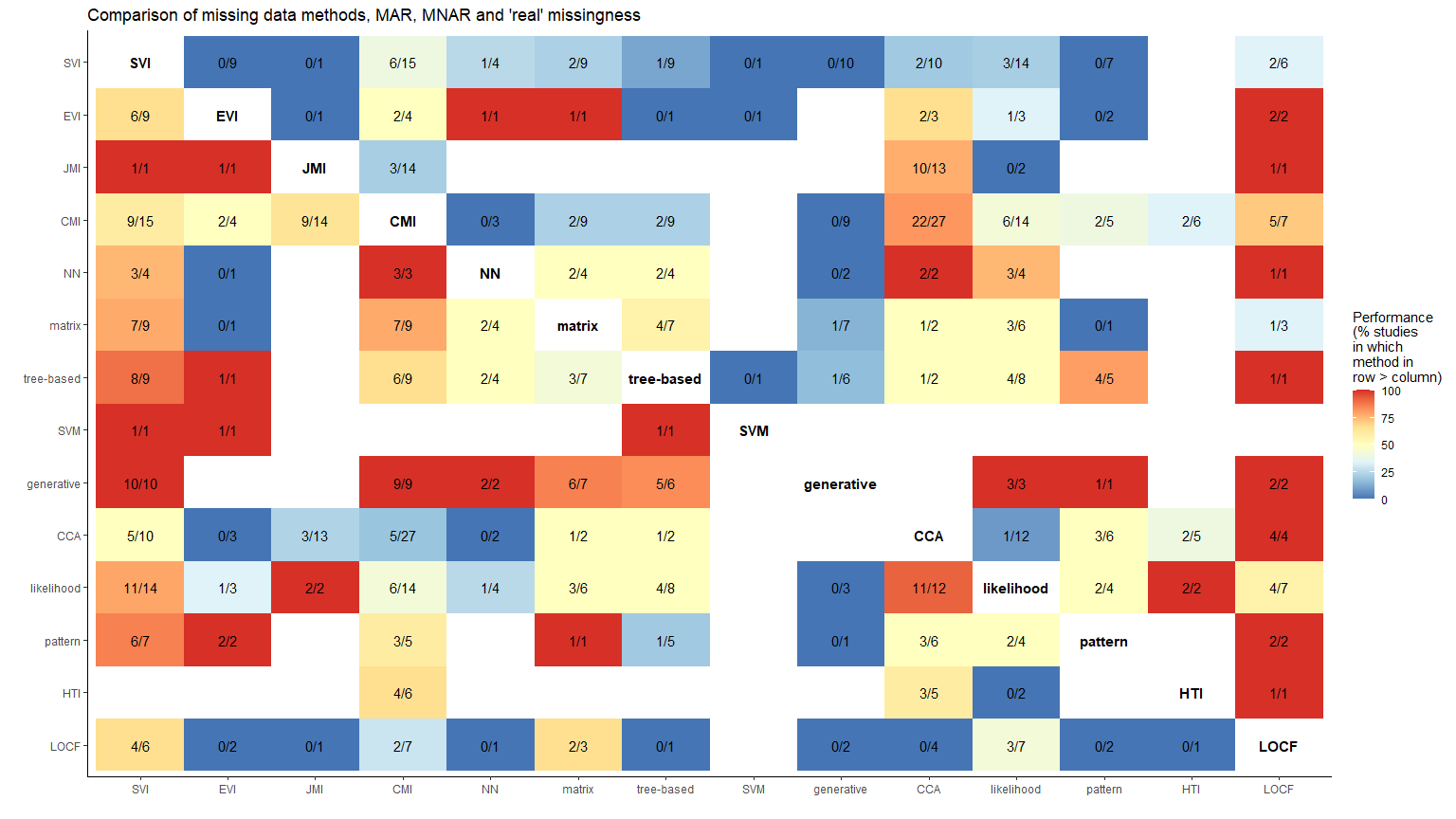


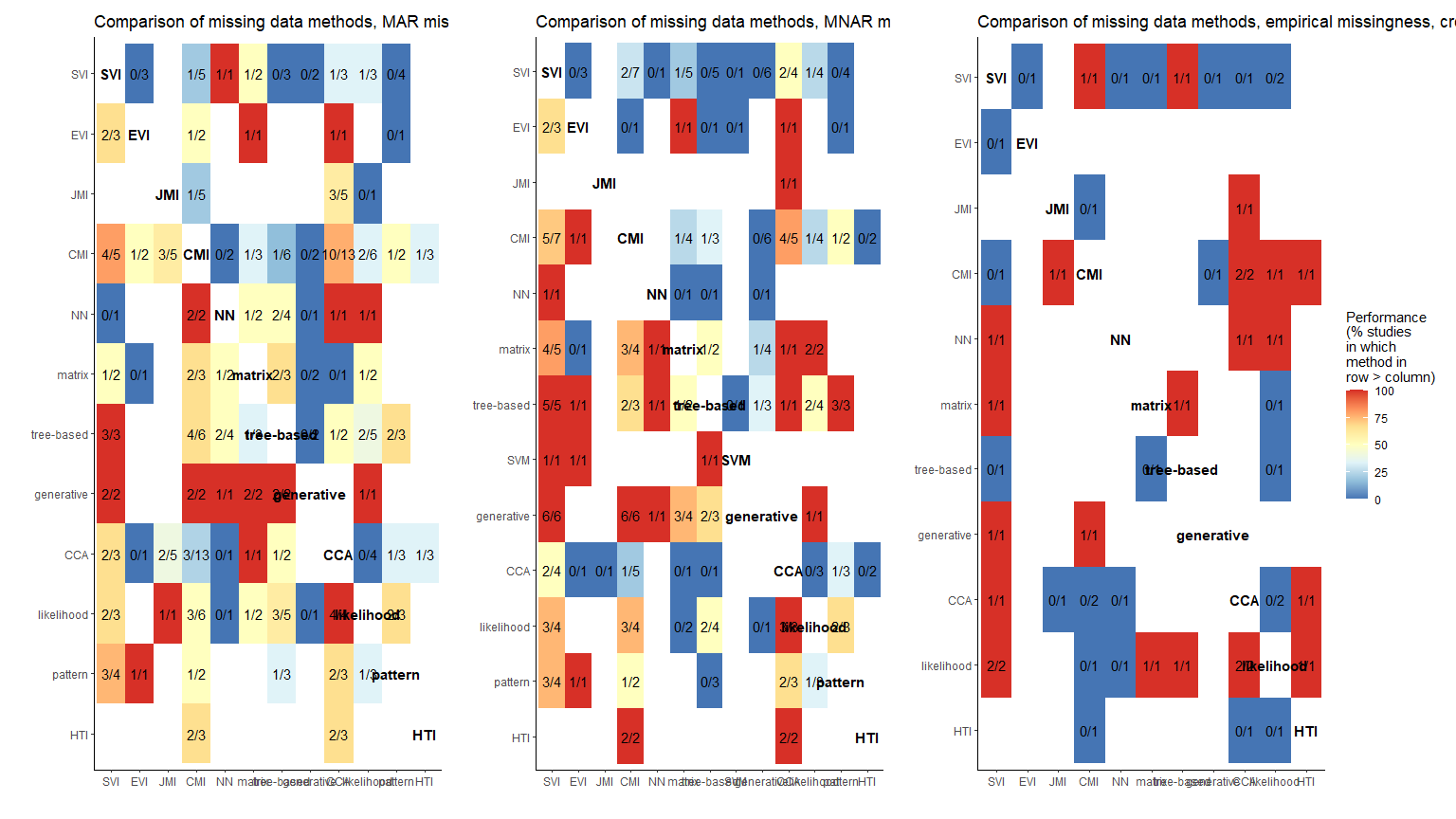
Figure. Flow chart of review process

# Results

* complete results presented in Figure:



* results for cross-sectional data



* results for longitudinal data

