

# Determinants of Cognitive Decline in a Large Cross-national Study Using Machine Learning

Anja K. Leist

University of Luxembourg, Institute for Research on Socio-Economic Inequality, Luxembourg

## THE SETTING

Dementia risk prediction scores that are currently available have only limited ability to predict risk on the individual level.

Risk scores with predictors that are easy and cost-effective to collect on a large scale are needed to

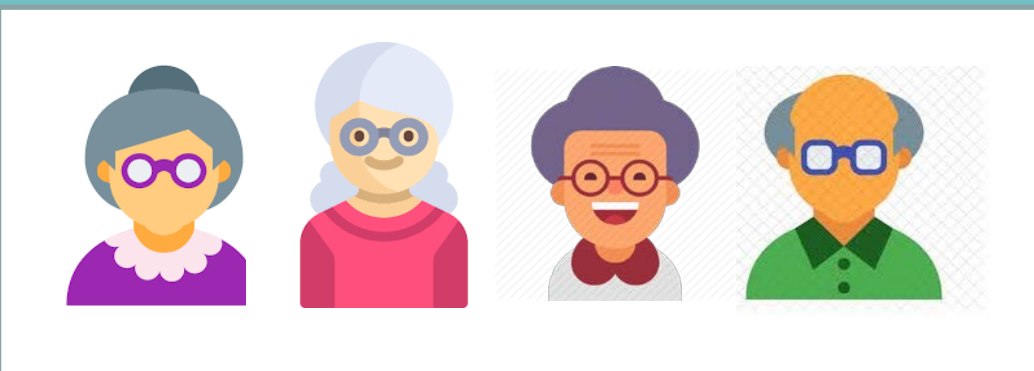
- Identify individuals at risk of cognitive decline and dementia
- Assign high-risk individuals to social-behavioral (multimodal) interventions

To improve dementia risk prediction, we need methods that can

- Process large number of predictors in many individuals over several follow-ups
- Detect non-linearities
- Detect highly complex non-additive relationships between variables

Recent advances in machine learning are able to process large numbers of risk and protective factors in cognitive aging more efficiently.

## PREDICTION WITH MACHINE LEARNING WHO WILL DECLINE?



## THIS STUDY

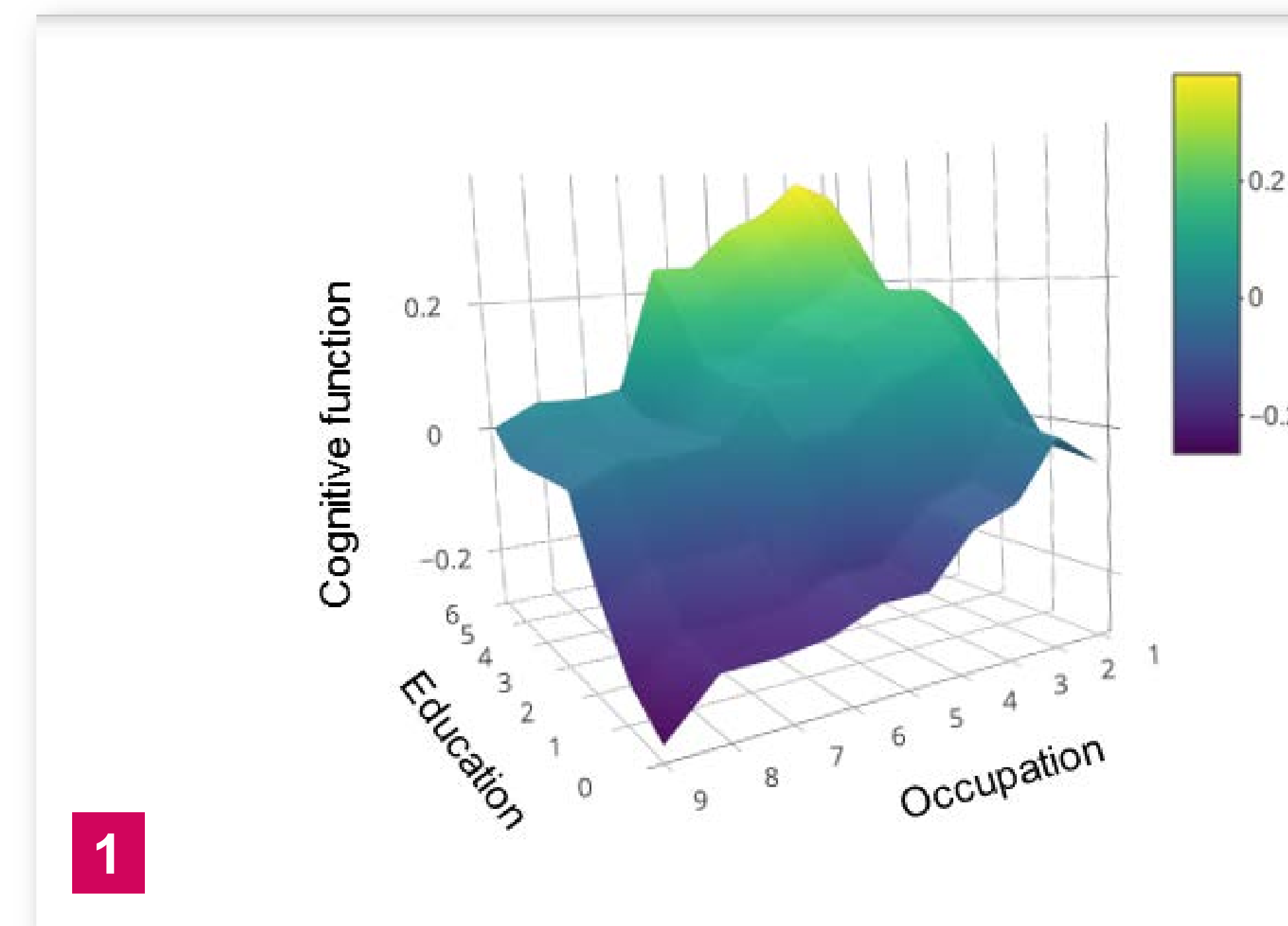
**Method.** A sample of 43,899 respondents from 13 European countries and Israel aged 50-101 years participated in at least two SHARE waves. We included an extensive set of variables related to demographics, health, functional limitations, skills, activities, childhood parental background, mental health, and ICT use to predict change in a composite score of immediate, delayed recall, and fluency over two years.

We chose Bayesian Additive Regression Trees (Chipman et al. 2010; Kapelner & Bleich 2014; Hill 2011; BARTmachine) and Gradient Boosting (Chen et al. 2016; xgboost) to investigate determinants of cognitive decline data-driven.

Models were built with training and out-of-sample test sets and four-fold cross-validation, with and without baseline adjustment. Variables were tested on importance, interactions, and inspected via partial dependency and individual conditional expectation plots.

**Results.** In fully adjusted models, variance explained in cognitive change was 30.2%. Cognitive decline was steeper after age 65, before levelling again off after age 72. Significant risk factors for cognitive decline (i.e. absence of practice effects) were educational attainment, maximum grip strength <30kg, decline in life satisfaction, and additional functional limitations.

## Non-Linear Interactions of Education and Occupation on Cognitive Functioning



1

## Visualization after Model Building

Figure 1. Non-linear Interactions with 3D Partial Dependency Plot

Figure 2. SHAP Plots (Lundberg, Erion, Lee 2018)

Figure 3. Years of Education; Partial Dependency Plot

## DATA SCIENCE TASKS: DESCRIPTION, PREDICTION, CAUSAL INFERENCE

Taken from Hernan, Hsu & Healy (2018):

*Data science* as umbrella term for all types of data analysis

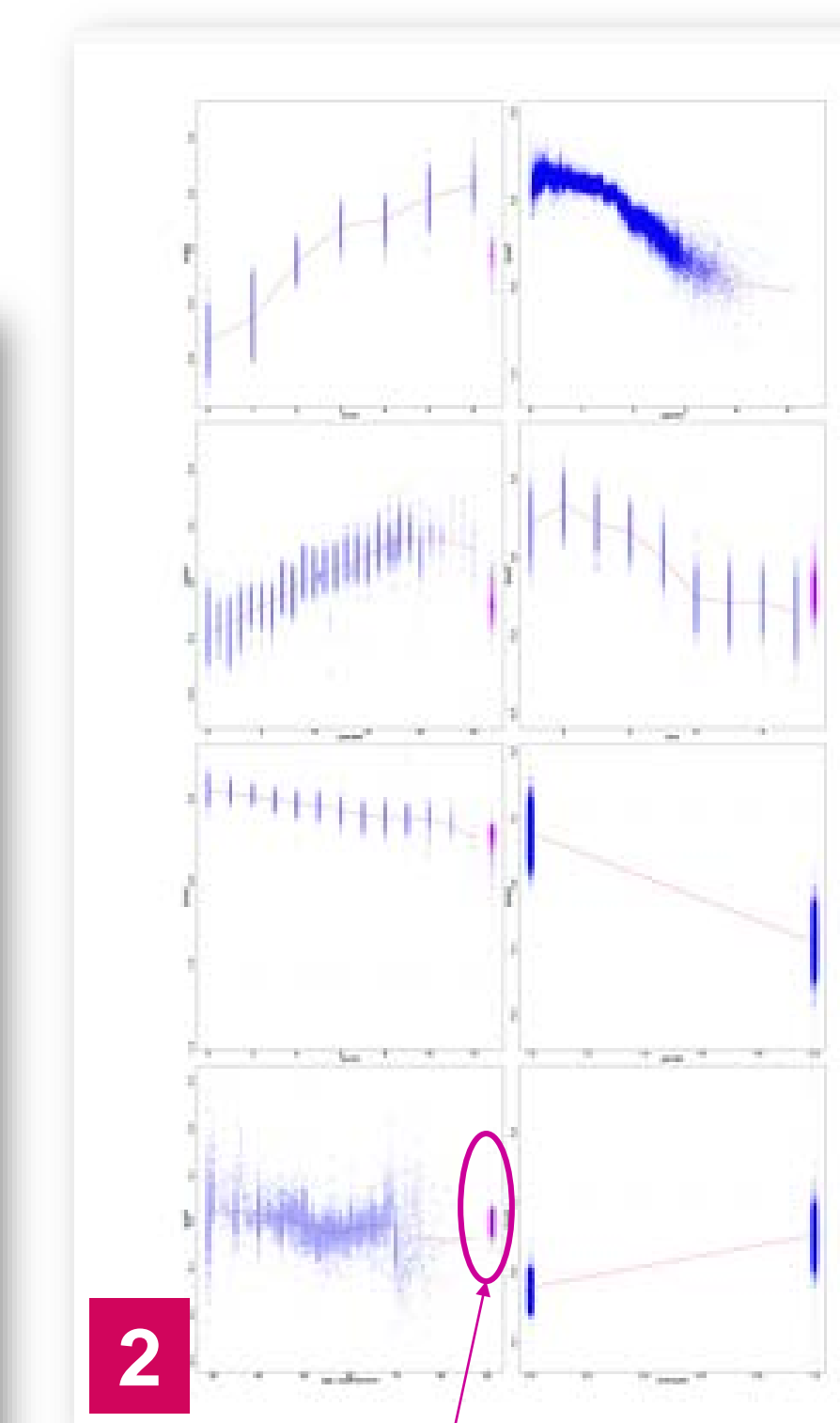
Scientific tasks:

- **Description:** Quantitative summary of characteristics or features, including **data visualization**
- **Prediction:** Mapping some characteristics or features to others, including associations and longitudinal complex predictions of e.g. health and disease outcomes
- **Causal inference:** Predicting certain features of the world if the world had been different → causal inference is counterfactual prediction. Choose confounders wisely or – if not possible – assess sensitivity to confounding

Key difference between prediction and causal inference: Causal inference needs domain-specific expert knowledge

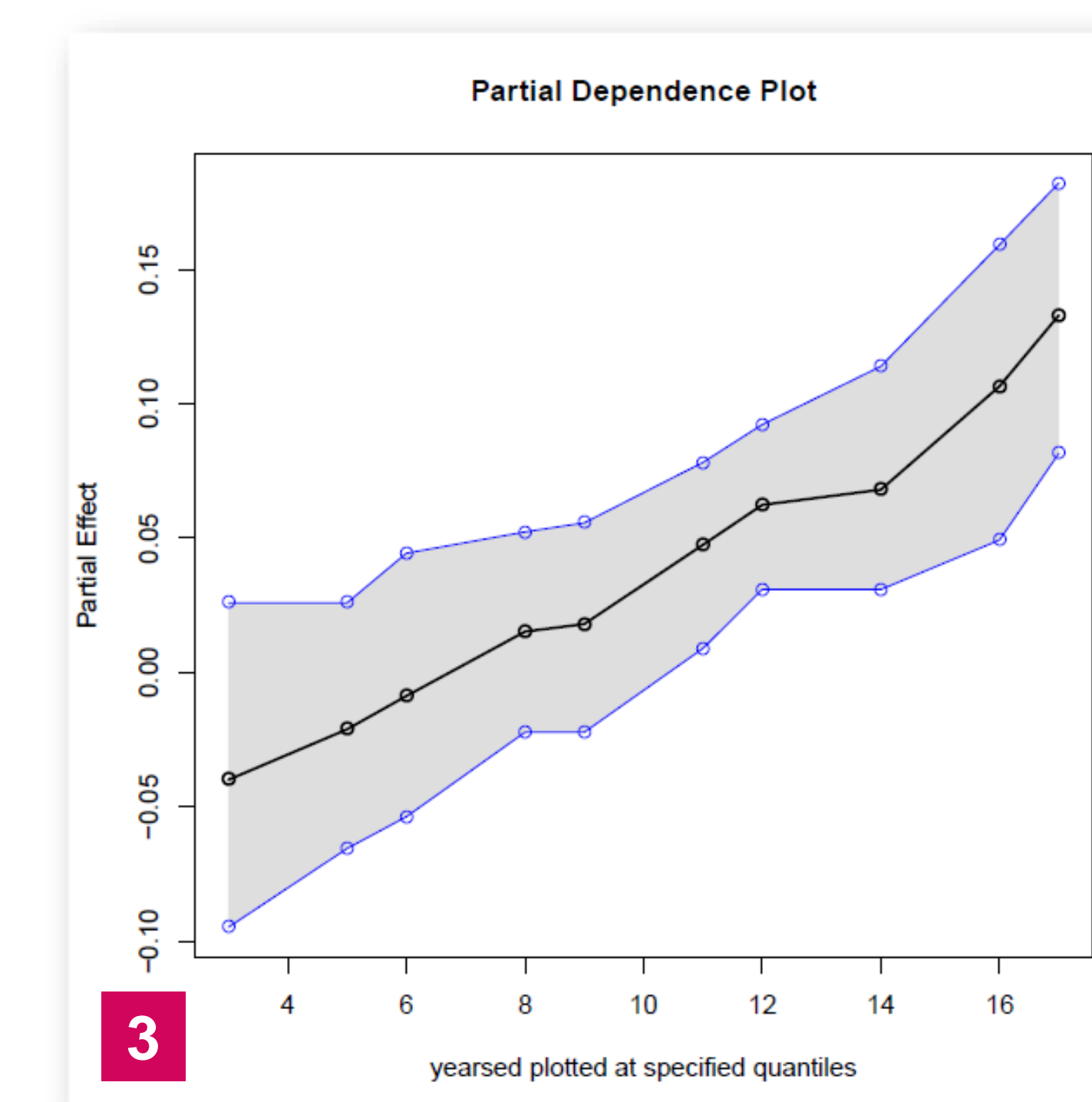
Characteristics of predictive models:

- No domain-specific expert knowledge necessary
- Use all available data for model building
- Model building can be automated, data-driven



2

Missing data profiles for each predictor



3

## LIMITS

Shortcomings of predictive models:

- Should be used with caution for assumptions about cause and effect: Selecting features may even increase bias in the model because the inclusion of certain features may condition on the wrong variables
- Should only be used for decision-making, interventions if guided by domain-specific expert knowledge: 'predictive' features are not necessarily the right features to intervene on

## CONCLUSIONS

- Predictive machine learning models can help to process large datasets and detect non-linear, possibly complex relationships. This study indicates that for social-behavioral research questions, strengths of machine learning lie in improved description and visualization, not necessarily predictive power.
- Specific for investigations of cognitive decline, long follow-ups are necessary. Model building with machine learning techniques still require heavy pre-processing of variables, domain-specific knowledge.
- Applying machine learning techniques with *causal inference* frameworks (Pearl 2009; Pearl & Mackenzie 2018) and the use of *causal directed acyclic graphs* seem more promising to advance our current scientific knowledge on cognitive aging and dementia.

## REFERENCES

- Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. arXiv:1603.02754v3
- Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1), 266-298.
- Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1), 217-240.
- Kapelner, A., & Bleich, J. (2014). bartMachine: Machine Learning with Bayesian Additive Regression Trees. arXiv:1312.2171v3
- Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. arXiv: 1802.03888
- Hernan, Hsu & Healy (2018). Data science is science's second chance to get causal inference right. A classification of data science tasks. arXiv:1804.10846
- Pearl, J. (2009). *Causality*. Cambridge University Press.
- Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The new science of cause and effect*. Basic Books.



Anja K. Leist, PhD

I am Associate Professor at the University of Luxembourg with a focus on health and cognitive functioning in older age from a life course and social epidemiological perspective. In 2018, I was awarded an ERC Starting Grant on the topic of 'Cognitive Aging: From Educational Opportunities to Individual Risk Profiles'. I am GSA Fellow and Steering Group member of the *World Young Leaders in Dementia* network (wyl dementia.org).

Contact: anja.leist@uni.lu