EMA
EUROPEAN MEDICINES AGENCY

Exploring the use of causal inference methods for clinical trials

BBS Seminar, October 2025

Florian Lasch

1. Motivating example: Initiation of symptomatic medication in Alzheimer’s Disease trials

2. Current challenges

3. Regulatory considerations

4. Key take aways

Disclaimer

The views expressed in this presentation and in the following panel discussion are the personal opinion of the author and should not be understood as being made on behalf of or reflecting the position of the EMA or any of its committees or working parties.

Causal inference for clinical trials

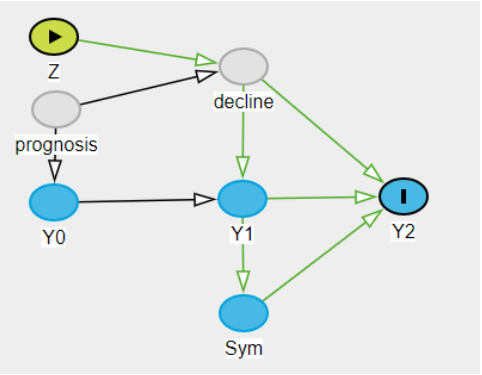
Estimand

Counterfactual notation

$$E[Y^{Z=1, ICE1=0} - Y^{Z=0, ICE1=0}]$$

Statistical estimand

Clarify assumptions: causal DAGs



Estimator

Explore violations of assumptions

Sensitivity estimator 1

Lasch et al. (2022)

Case study:

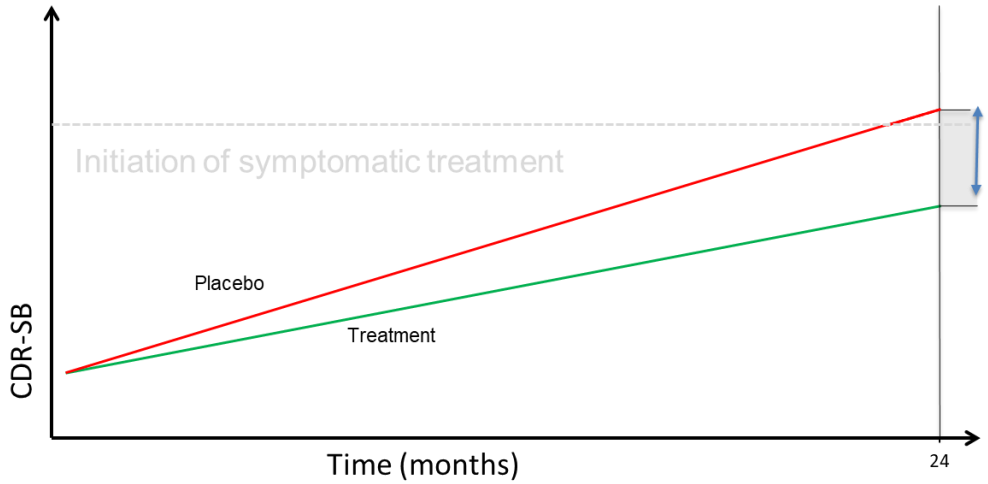
Symptomatic treatment in trials for Alzheimer’s Disease

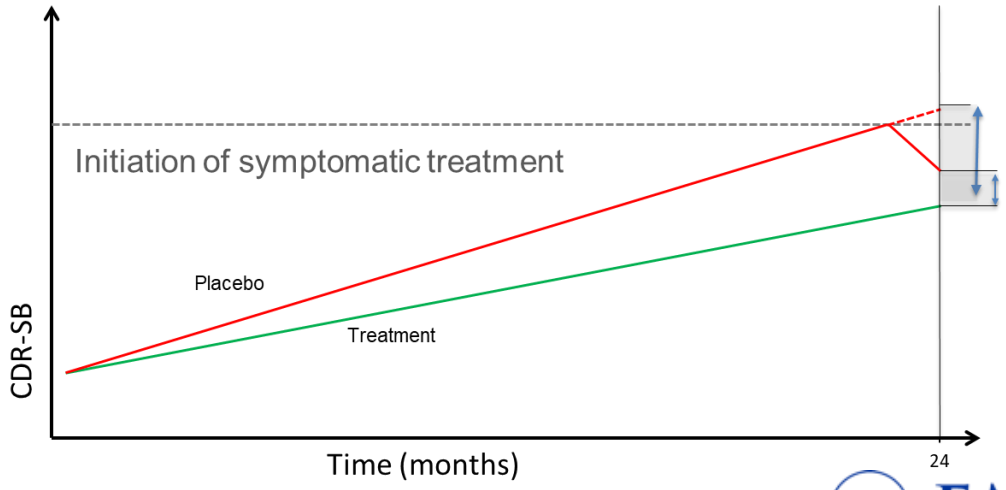
Lasch et al. (2022): A Simulation Study on the Estimation of the Effect in the Hypothetical Scenario of No Use of Symptomatic Treatment in Trials for Disease-Modifying Agents for Alzheimer’s Disease

Lasch et al. (2025): Full article: Comparison of g-estimation approaches for handling symptomatic medication at multiple timepoints in Alzheimer's Disease with a hypothetical strategy

Symptomatic treatment in Alzheimer's Disease trials

- **Estimand**
- Population: Patients with prodromal AD
- Treatment: Disease modifying treatment vs Placebo
- Endpoint: CDR-SB at 24 months
- Summary measure: Difference in mean CDR-SB between treatment arms
- **Intercurrent event: Initiation of symptomatic treatment**





EMA

Symptomatic treatment in Alzheimer's Disease trials

EMA Guideline on Clinical investigation of medicines for the treatment of Alzheimer's disease (2018):

"[…], providing that reliable methods of estimation can be identified, an appropriate target of estimation could be based on a **hypothetical scenario** in which the new concomitant medication or modifications in the dose of concomitant medications had not been introduced."

→ **What is a reliable method of estimation?**

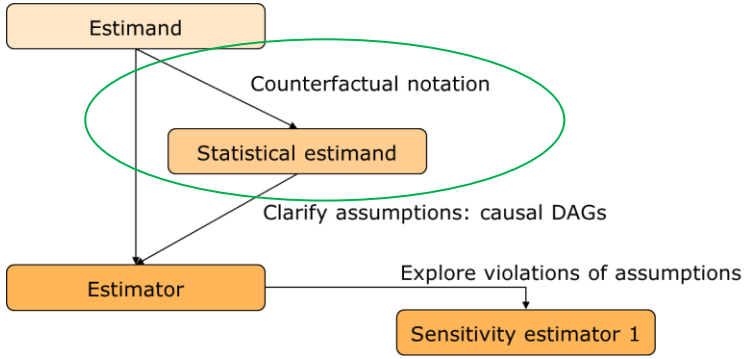
- Common practice:
- (i) set values after the initiation of symptomatic treatment as missing
- (ii) apply missing data approaches using mixed models for repeated measures, Inverse probability weighting, etc.

Alzheimer's Disease – statistical estimand

Statistical estimand:

$$E\left[\ Y_2^{Z=1,Sym=0} - Y_2^{Z=0,Sym=0}\ \right]$$

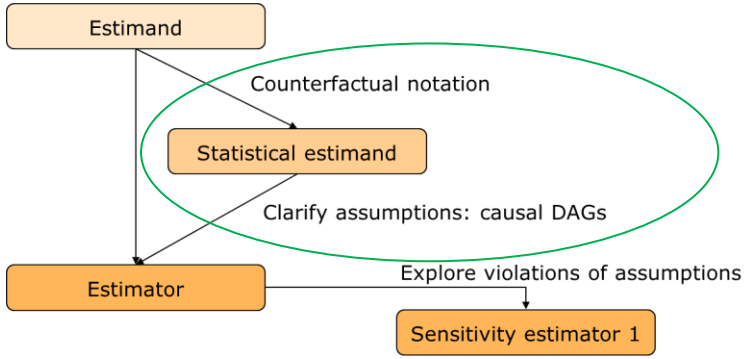
- Z: randomized treatment
- Sym: Initiation of symptomatic treatment
- Y_t: observed outcome at time t

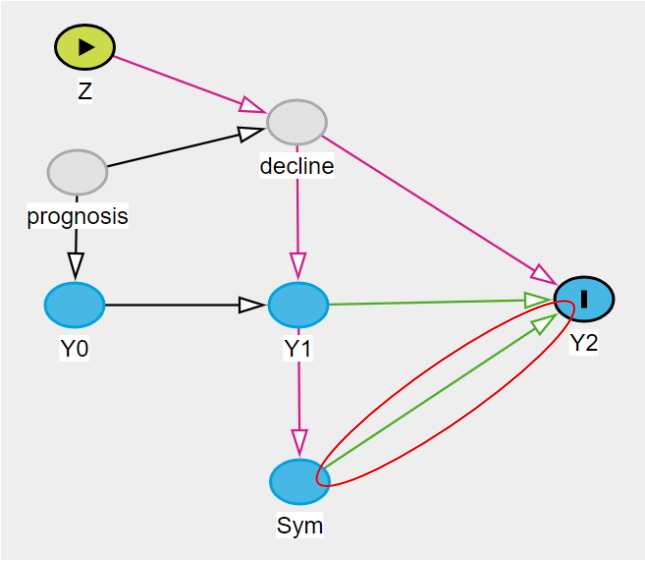


Alzheimer's Disease – causal structure

Statistical estimand:

$$E\left[\ Y_2^{Z=1,Sym=0} - Y_2^{Z=0,Sym=0}\ \right]$$





Observation based on the assumed causal structure

- Sym is a mediator of the effect of Z on Y_2
- The estimand of interest corresponds to the controlled direct effect of Z on Y_2
- de-mediation approaches like g-estimation are a candidate for the estimation

Alzheimer's Disease – g-estimation

- **G-estimation***

1. Estimate the effect of the mediator Sym on Y_2

   Predict the probability of Sym: $P(Sym = 1) \sim Y_1 + Y_0 + Z$

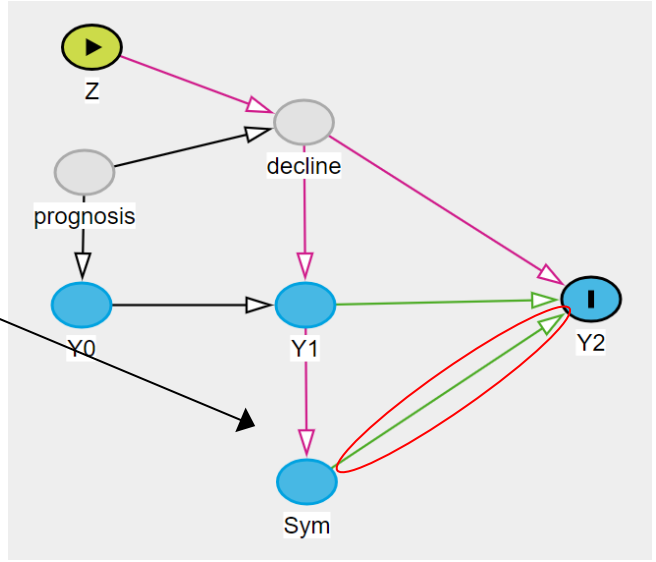
   Estimate the effect of Sym: $Y_2 \sim Z + Y_0 + Y_1 + p_{sym} + Sym$

2. De-mediate the effect of Sym from Y_2

   $R_2 = Y_2 - Sym * \beta_{sym}$

3. Estimate the effect of Z on the de-mediated values

   $R_2 \sim Z + Y_0$



Z: randomized treatment

Sym: Initiation of symptomatic treatment

Y_t: observed CDR-SB at time t

*: Loh et al.: Estimation of Controlled Direct Effects in Longitudinal Mediation Analyses with Latent Variables in Randomized Studies

EMA

Alzheimer's Disease - Simulation study

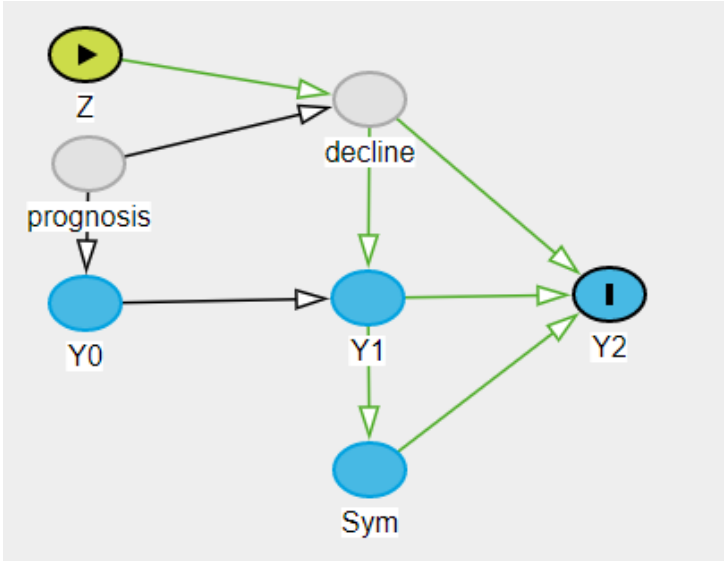
- **Objectives**:

- Quantify the performance of commonly used estimators (bias, T1E / power)

- Compare the performance to de-mediation via g-estimation

- **Data generating mechanism**

- Disease progression model: beta regression model with Richard's logistic link function g



- $\frac{Y^*_{t,i}}{18} \sim \beta(a,b), \quad g(x) = \left(\frac{x^\beta}{1-x^\beta}\right)^{\frac{1}{\beta}}$

- für $\tilde{t} > t$: $g(\overline{Y^*_{\tilde{t},i}}) = g(Y^*_{t,i}) + \alpha_i * \frac{\tilde{t}-t}{52} * (E_{DM})^{treat_i} * (E_C)^{confound_i}$

- α_i: random decline rate

→ Linearity Assumptions about the functional form are violated for all methods

EMA

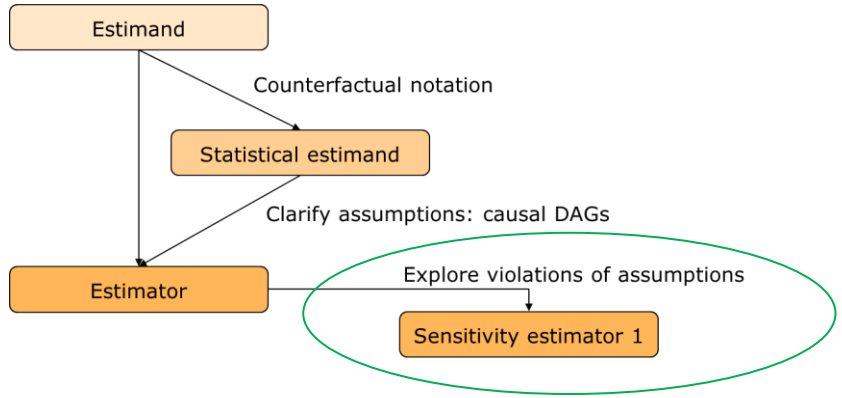
Simulation results – alternative hypothesis

Estimator	Bias (SE*)	Coverage (SE*)	sqrMSE (MSE (SE*))	CI length (SE*)*	empirical power (95% CI*)
reference estimator – linear model using Y_2^*	0.004 (0.003)	94.61 (0.945)	0.235 (0.088 (0.001))	1.167 (0.001)	**0.896 (0.89, 0.902)**
Observed Value	0.035 (0.003)	94.24 (0.941)	0.229 (0.084 (0.001))	1.13 (0.001)	**0.895 (0.889, 0.901)**
Mixed Effects Model with observed values	0.037 (0.003)	94.79 (0.946)	0.261 (0.108 (0.002))	1.281 (0.001)	**0.804 (0.796, 0.812)**
Observed Values - adjusted	0.122 (0.003)	92.73 (0.924)	0.24 (0.091 (0.001))	1.088 (0.001)	**0.855 (0.848, 0.862)**
Loh's g-estimation, model-based SE	0.007 (0.003)	94.45 (0.943)	0.235 (0.088 (0.001))	1.163 (0.001)	**0.897 (0.891, 0.903)**
Loh's g-estimation, Bootstrap based SE	0.007 (0.003)	94.38 (0.942)	0.235 (0.088 (0.001))	1.162 (0.001)	**0.893 (0.887, 0.899)**
Linear Sequential g-Estimation	0.014 (0.003)	94.27 (0.941)	0.233 (0.087 (0.001))	1.148 (0.001)	**0.897 (0.891, 0.903)**
Predictive Mean Matching (PMM)	0.152 (0.003)	90.41 (0.899)	0.264 (0.107 (0.001))	1.121 (0.001)	**0.8 (0.793, 0.808)**
PMM worsening adjustment 0.5	0.127 (0.003)	91.85 (0.915)	0.26 (0.105 (0.001))	1.154 (0.001)	**0.805 (0.797, 0.813)**
PMM worsening adjustment 2	0.051 (0.003)	94.12 (0.939)	0.266 (0.112 (0.002))	1.288 (0.001)	**0.783 (0.775, 0.791)**
PMM worsening adjustment 3	0 (0.004)	94.57 (0.944)	0.285 (0.129 (0.002))	1.402 (0.001)	**0.756 (0.748, 0.765)**
Inverse Probability WeightingWeighting	0.118 (0.004)	91.57 (0.912)	0.332 (0.175 (0.003))	1.532 (0.004)	**0.593 (0.583, 0.603)**
doubly robust Inverse Probability Weighting	0.028 (0.003)	94.06 (0.939)	0.252 (0.101 (0.001))	1.217 (0.001)	**0.855 (0.848, 0.862)**
Mixed Effects Model	0.107 (0.003)	93.08 (0.928)	0.275 (0.119 (0.002))	1.27 (0.001)	**0.746 (0.738, 0.755)**

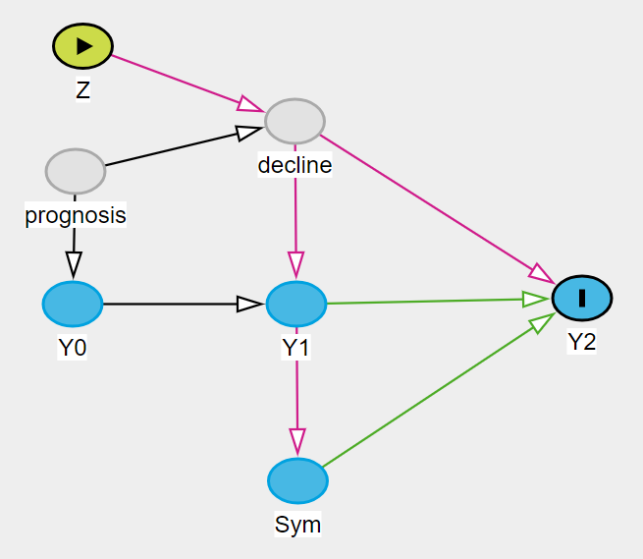
*: Monte Carlo estimates of the standard errors. Performance parameter are rounded to three digits.

Sensitivity analysis

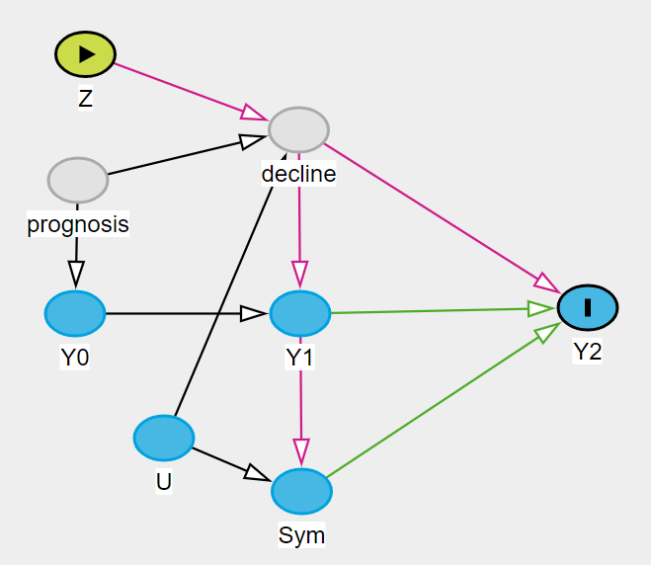
- Different causal structure
- Violation to assumed DGM (linearity…)



Without unmeasured confounder



With unmeasured confounder



Simulation results – different causal structure

Estimator	Bias (SE*)	Coverage (SE*)	sqrMSE (MSE (SE*))	CI length (SE*)	empirical power (95% CI*)
reference estimator – linear model using Y_2^*	0 (0.003)	94.83 (0.947)	0.214 (0.072 (0.001))	1.048 (0.001)	**0.835 (0.828, 0.842)**
Observed Value	0.027 (0.003)	94.7 (0.946)	0.208 (0.068 (0.001))	1.019 (0.001)	**0.831 (0.823, 0.838)**
Mixed Effects Model with observed values	0.028 (0.003)	94.74 (0.946)	0.234 (0.086 (0.001))	1.145 (0.001)	**0.729 (0.72, 0.738)**
Observed Values - adjusted	0.071 (0.003)	93.84 (0.936)	0.21 (0.069 (0.001))	1.003 (0.001)	**0.797 (0.789, 0.805)**
Loh's g-estimation, model-based SE	-0.002 (0.003)	94.86 (0.947)	0.214 (0.072 (0.001))	1.051 (0.001)	**0.835 (0.828, 0.842)**
Loh's g-estimation, Bootstrap based SE	-0.002 (0.003)	94.84 (0.947)	0.214 (0.072 (0.001))	1.05 (0.001)	**0.831 (0.824, 0.839)**
Linear Sequential g-Estimation	0.002 (0.003)	94.77 (0.946)	0.213 (0.072 (0.001))	1.041 (0.001)	**0.836 (0.828, 0.843)**
Predictive Mean Matching (PMM)	0.122 (0.003)	91.54 (0.911)	0.236 (0.085 (0.001))	1.032 (0.001)	**0.692 (0.683, 0.701)**
PMM worsening adjustment 0.5	0.1 (0.003)	92.69 (0.924)	0.234 (0.085 (0.001))	1.063 (0.001)	**0.694 (0.685, 0.703)**
PMM worsening adjustment 2	0.032 (0.003)	94.83 (0.947)	0.249 (0.097 (0.001))	1.205 (0.001)	**0.674 (0.665, 0.683)**
PMM worsening adjustment 3	-0.013 (0.003)	94.89 (0.948)	0.273 (0.117 (0.002))	1.331 (0.001)	**0.636 (0.626, 0.645)**
Inverse Probability Weighting	0.078 (0.004)	92.3 (0.92)	0.305 (0.149 (0.002))	1.431 (0.003)	**0.512 (0.502, 0.521)**
doubly robust Inverse Probability Weighting	0.014 (0.003)	94.13 (0.939)	0.233 (0.085 (0.001))	1.119 (0.001)	**0.767 (0.758, 0.775)**
Mixed Effects Model	0.087 (0.003)	93.74 (0.935)	0.245 (0.094 (0.001))	1.144 (0.0008)	**0.658 (0.649, 0.667)**

***: Monte Carlo estimates of the standard errors**

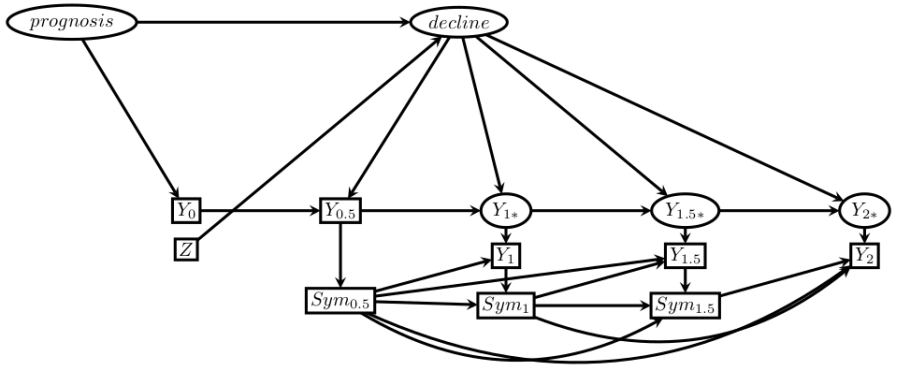
Performance parameter are rounded to three digits.

EMA

Alzheimer's Disease – multiple timepoints

Comparison of g-estimation approaches for handling symptomatic medication at multiple timepoints in Alzheimer's Disease with a hypothetical strategy

Lasch Florian[1,2,*,#], Guizzaro Lorenzo[1,3,*], Loh Wen Wei[4]



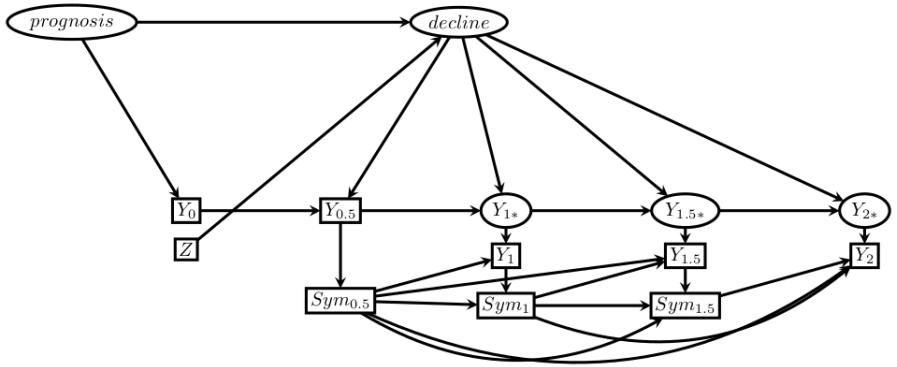
Extension of de-mediation approach

- If the effects of Sym are the same at each timepoint, the precision of the estimates of the effect of the symptomatic treatment can be increased
- Options:
  - Average the effects $Sym_t \rightarrow Y_{t+0.5}$
  - Average the effects $Sym_t \rightarrow Y_2$
  - Iteratively average the effects $Sym_t \rightarrow Y_2$ until convergence

Alzheimer's Disease – multiple timepoints

Comparison of g-estimation approaches for handling symptomatic medication at multiple timepoints in Alzheimer's Disease with a hypothetical strategy

Lasch Florian[1,2,*,#], Guizzaro Lorenzo[1,3,*], Loh Wen Wei[4]



	Empirical Bias*	Empirical standard deviation of the effect estimate	Mean model derived Standard Error	Mean Bootstrap based Standard Error	Empirical Power based on bootstrapping	Mean Jackknife based Standard Error	Empirical Power based on jackknife
MMRM	0.858	2.301	2.275	NA	0.58[#]	NA	NA
Established g-estimation approach for longitudinal de-mediation	0.115	1.988	1.818	1.981	0.8523	2.146	0.7973
Modification 1 Averaging the de-mediation on Y_t	0.052	1.825	1.800	1.818	0.8929	1.859	0.8871
Modification 2 Averaging the de-mediation effects on Y_2	0.059	1.912	1.806	1.888	0.8851	2.003	0.8538
Modification 3 Iterative averaging the de-mediation effects on Y_2	0.056	1.901	1.805	1.882	0.8859	1.989	0.8585

* The expected value of the effect (active treatment – placebo) under the alternative hypothesis is -5.83 on the ADAS-Cog 13 scale, where lower values correspond to a less sever disease. The bias is calculated as model estimate – expected value. A positive bias corresponds to an underestimation of the effect.

[#]: model-based power is reported for the observed values estimator and the MMRM as the model-derived standard errors are unbiased.

Assumption – structural mean model

The role of post intercurrent event data in the estimation of hypothetical estimands in clinical trials

Jonathan W. Bartlett[1] and Rhian M. Daniel[2]

- Use of post-intercurrent event data valid only under the assumption that there is no interaction between treatment, occurrence of the intercurrent event or covariates
- Simulation study shows bias of de-mediation approaches, if interaction is ignored
- Already highlighted by Loh et al*:
  - Structural mean model needs to be specified correctly (including interactions)

*: Loh et al.: Estimation of Controlled Direct Effects in Longitudinal Mediation Analyses with Latent Variables in Randomized Studies

Open questions

- Handling of missing data
- Mis-specification of Structural Mean Model
  - Bartlett & Daniels: importance of specifying the mean model correctly, including the interactions
- Balancing robustness advantages with plausibility of additional assumptions

Regulatory considerations

- Only relevant for estimating a hypothetical strategy?
- Is the Alzheimer's Disease Guideline an anomaly?
  - Treatment switching in oncology
  - Other examples of hypothetical strategies in EMA guidelines:
    - Depression – use of alternative anti-depressants
    - Bipolar disease – use of alternative treatment
    - Diabetes – rescue / additional medication

Regulatory considerations

- Only relevant for estimating a hypothetical strategy?
- Is the Alzheimer's Disease Guideline an anomaly?
  - Treatment switching in oncology
  - Other examples of hypothetical strategies in EMA guidelines:
    - Depression – use of alternative anti-depressants
    - Bipolar disease – use of alternative treatment
    - Diabetes – rescue / additional medication

- Possible conflicts with current guidelines
  - Caution: use of post-baseline variables can introduce confounding, depends on causal structure and DGM (EMA Guideline on adjustment for baseline covariates in clinical trials)

Opportunities and challenges

- Causal inference thinking can already help NOW
  - Causal DAGs make assumptions explicit
  - Causal DAGs facilitate the identification of relevant sensitivity analysis

- Challenges of applying causal inference methods to clinical trials
  - Performance of methods only established via simulation study?
  - Are (additional) assumptions plausible?
  - More thinking needed!

Key take-aways

- In most situations, established statistical methods (+randomisation) for analysing clinical trials are fit-for-purpose (Occam's razor!)
- In some situations, more complex methods might be needed for unbiased estimation or to increase efficiency (hypothetical strategy)
- The consistent application of the estimands framework can help to identify these settings
- Causal inference != only statistical methods

EMA

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Thank you

Florian.lasch@ema.europa.eu

Follow us