Using Longitudinal Targeted Maximum Likelihood Estimation in Complex Settings with Dynamic Interventions

Michael Schomaker, Miguel Angel Luque-Fernandez, Valeriane Leroy, Mary-Ann Davies

European Causal Inference Meeting, Bremen, Germany

27 March 2019

Motivating Question (I)

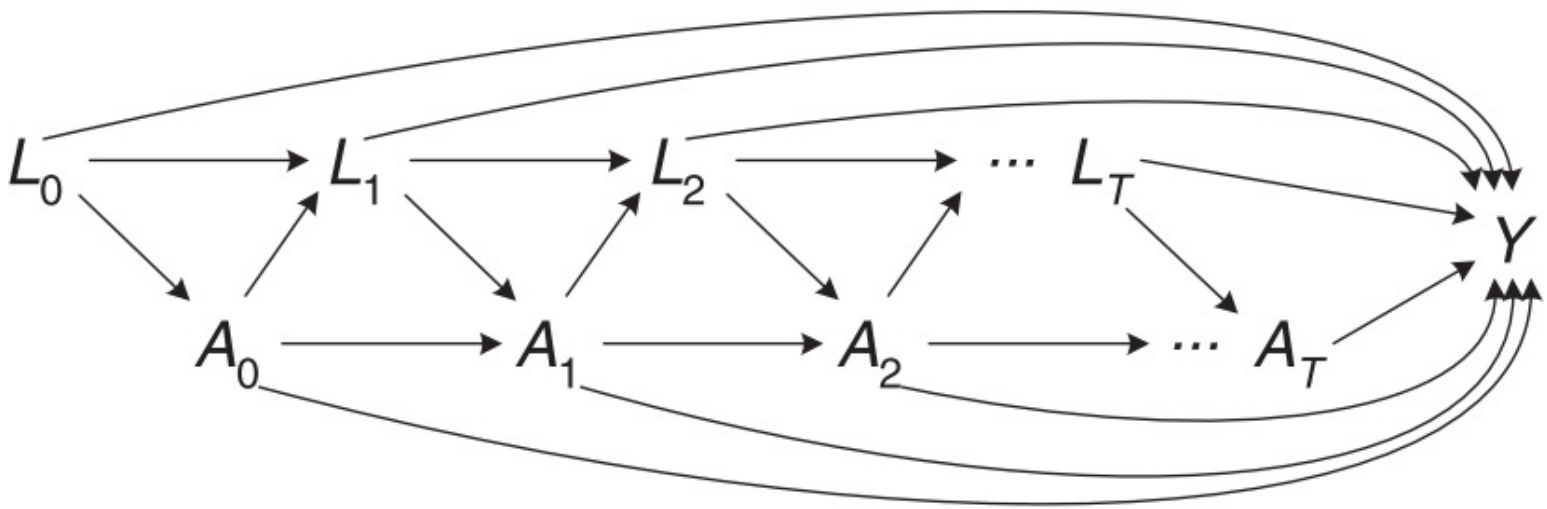
- Antiretroviral treatment (ART) is known to be highly effective
- Treatment initiation is still often delayed

  (former guidelines; concerns about toxicities, non-adherence, drug resistance; logistical challenges; cost considerations)
- There is limited knowledge about the optimal timing of antiretroviral treatment initiation in children and adolescents
- It is no longer ethically possible to conduct a trial
- Regular update of treatment guidelines by WHO

Motivating Question (II)

Of interest: the effect of different treatment initiation rules on mortality and growth

Time-dependent confounding affected by prior treatment:



Source: Daniel et al. [1]

Notation

- Follow-up time: $t = 0, 1, \ldots, T$ months
- Outcome: $Y_t =$ HAZ (height-for-age z-score)
- Intervention: $A_t =$ antiretroviral therapy (ART)
- Confounders:
  - Time-Varying: $\mathbf{L}_t =$ CD4 count, CD4%, WAZ (=WHO stage)
  - Baseline: $\mathbf{L}_0 =$ CD4 count, CD4%, WAZ, HAZ, sex, age, year, region
- Censoring: C_t
- Survival: $S_t =$ Death
- History: e.g. $\bar{A}_t = (A_0, \ldots, A_t)$
- **Counterfactual:** e.g. $Y_t^{\bar{a}_t}$
- Intervention rule: e.g. $d_t(\bar{\mathbf{L}}_t)$ [assigns A_t as a function of $\bar{\mathbf{L}}_t$]

Target Quantity

$\psi_T = \mathbb{E}(Y_T^{\bar{d}})$

i.e. the expected value of Y at time T under an intervention rule $\bar{d}$ which assigns A_t as a function of $\bar{\mathbf{L}}_t$ and sets C_t and S_t deterministically to 1.

$\rightarrow$ In the example, we consider

$$\psi_{30} = \mathbb{E}(Y_{30}^{\bar{d}^j})$$

i.e. the mean HAZ at 30 months, under no censoring, for a given treatment rule $\bar{d}^j$ to be the quantity of interest

Interventions (Static and Dynamic)

With $L^1 = \text{CD4 count}$ and $L^2 = \text{CD4\%}$ we evaluate:

$$\bar{d}_t^1 = \{c_t = 1; s_t = 1; a_t = 1 \quad \text{for} \quad \forall t$$

$$\bar{d}_t^2(\overline{L^1}_t, \overline{L^2}_t) = \begin{cases} c_t = 1; s_t = 1; \quad a_t = 1 & \text{if} \quad L^1{}_t^{\bar{d}} < 750 \quad \text{or} \quad L^2{}_t^{\bar{d}} < 25\% \\ & \quad \text{or} \quad a_{t-1} = 1 \\ c_t = 1; s_t = 1; \quad a_t = 0 & \text{otherwise} \end{cases}$$

$$\bar{d}_t^3(\overline{L^1}_t, \overline{L^2}_t) = \begin{cases} c_t = 1; s_t = 1; \quad a_t = 1 & \text{if} \quad L^1{}_t^{\bar{d}} < 350 \quad \text{or} \quad L^2{}_t^{\bar{d}} < 15\% \\ & \quad \text{or} \quad a_{t-1} = 1 \\ c_t = 1; s_t = 1; \quad a_t = 0 & \text{otherwise} \end{cases}$$

$$\bar{d}_t^4 = \{c_t = 1; s_t = 1; a_t = 0 \quad \text{for} \quad \forall t$$

Data Example: Comparison of 3 Estimators

Estimate $\psi_{30} = \mathbb{E}(Y_{30}^{\bar{d}^j_{30}})$ for different interventions:

(i) g-formula, manual implementation which includes prior clinical knowledge using additive regression models

(ii) LTMLE, manual implementation which includes prior clinical prior knowledge using additive regression models

(iii) LTMLE, "automated" (using `ltmle`), using a data-adaptive approach (super learning with 6 "simple learners", computational constraints)

"Prior clinical knowledge": children who are sicker at presentation will have a different disease trajectory from patients who are healthier at presentation (non-linear interactions)

Methodological Motivation for Comparison

- doubly robust estimators rarely applied under long follow-up, gradually declining sample size, dynamic interventions, and multiple time-dependent confounders

- no detailed comparison for complex longitudinal data between (parametric) g-formula and LTMLE yet

Also: simulations for different LTMLE estimation approaches under realistic, challenging settings (as above) may be informative

Estimator I: LTMLE

Using the iterative conditional expectation rule and the assumptions of positivity, consistency and sequential conditional exchangability[1], one can show[2] that

(for $\mathbf{L}_t \to Y_t \to A_t \to C_t \to S_t$, $Y_t \in \mathbf{L}_t$ for $t < T$ and $\mathbf{A}_t = \{A_t, C_t, S_t\}$)

$$\mathbb{E}(Y_T^{\bar{d}}) = \mathbb{E}(\mathbb{E}(\ldots\mathbb{E}(\mathbb{E}(Y_T|\bar{\mathbf{A}}_{T-1} = \bar{d}_{T-1}, \bar{\mathbf{L}}_T)|\bar{\mathbf{A}}_{T-2} = \bar{d}_{T-2}, \bar{\mathbf{L}}_{T-1})\ldots|\bar{A}_0 = d_0, \mathbf{L}_0)|\mathbf{L}_0).$$

The LTMLE estimator (van der Laan and Gruber, *IJB*, 2012 [3]) is based on the above equality.

A *targeted* step for each t enables doubly robust inference with respect to ψ_T, the quantity of interest.

[1] let's assume the assumptions are met for now

[2] Bang and Robins, *Biometrics*, 2005 [2]

Algorithm (I)

For $t = T, ..., 1$:

1. Use an appropriate regression model to estimate $\mathbb{E}(Y_t|\bar{\mathbf{A}}_{t-1}, \bar{\mathbf{L}}_t)$. The model is fitted on all subjects that are uncensored and alive (until $t-1$).

   Note that the outcome refers to the measured outcome for $t = T$ and to the prediction (of the conditional outcome) from step 3d (of iteration $t-1$) if $t < T$.

2. Now, plug in $\bar{\mathbf{A}}_{t-1} = \bar{d}_{t-1}$ based on rule $\bar{d}$ and use the regression model from step 1 to predict the outcome at time t, i.e. $\tilde{Y}_t^{\bar{d}}$.

Algorithm (II)

3. To improve inference with respect to ψ_t update the initial estimate of step 2:

   a) the outcome refers again to the measured outcome for $t = T$ and to the prediction from item 3d (of iteration $t - 1$) if $t < T$.

   b) the offset is the original predicted outcome from step 2 (iteration t).

   c) the estimated "clever covariate" refers to the cumulative product of inverse treatment and censoring probabilities:

$$\hat{H}(\bar{A}, \bar{C}, \bar{\mathbf{L}})_{t-1} = \prod_{s=0}^{t-1} \frac{I(\bar{\mathbf{A}}_s = \bar{d}_s) \times I(\bar{C}_s = 1)}{\begin{array}{r}\hat{\mathbb{P}}(\mathbf{A}_s = \bar{d}_s | \bar{\mathbf{L}}_s = \bar{\mathbf{l}}_s, \bar{\mathbf{A}}_{s-1} = \bar{d}_{s-1}, \bar{C}_{s-1} = 1) \\ \times \hat{\mathbb{P}}(C_s = 1 | \bar{\mathbf{L}}_s = \bar{\mathbf{l}}_s, \bar{\mathbf{A}}_{s-1} = \bar{d}_{s-1}, \bar{C}_{s-1} = 1)\end{array}}$$

   d) Predict the (updated) outcome, $\tilde{Y}_t^{\bar{d}}$, based on the model defined through 3a, 3b, and 3c.

Algorithm (III)

For $t = 1$:

4. The estimate $\hat{\psi}_T$ is obtained by calculating the mean of the predicted outcome from step 3d (where $t = 1$).

5. Confidence intervals can, for example, be obtained using the vector of the estimated influence curve of ψ_T, which can be written as

$$\widehat{\mathrm{IC}}(\psi_T) = \left\{ \sum_{s=1}^{T} \hat{H}(\bar{A}, \bar{C}, \bar{\mathbf{L}})_{s-1} \left[\tilde{Y}_s^{\bar{d}} - \tilde{Y}_{s-1}^{\bar{d}} \right] \right\} + \tilde{Y}_1^{\bar{d}} - \hat{\psi}_{T,\mathrm{TMLE}}$$

An asymptotically normal 95% confidence interval is then given by

$$\left[\hat{\psi}_{\mathrm{TMLE}} \pm 1.959964 \sqrt{\widehat{\mathrm{Var}}(\widehat{\mathrm{IC}})/n} \right] .$$

Estimator II: the g-formula[3] (only brief idea)

Here, with $\bar{\mathbf{A}}_t = \{\bar{A}_t, \bar{C}_t, \bar{S}_t\}$, $\mathbf{L}_t \to Y_t \to A_t \to C_t \to S_t$, $Y_t \in \mathbf{L}_t$ ($t < T$), we can write

$$\psi_T = \quad \mathbb{E}(Y_T^{\bar{d}}) = \int_{\bar{\mathbf{l}} \in \bar{\mathbf{L}}_t} \left\{ \begin{array}{l} \mathbb{E}(Y_T | \bar{\mathbf{A}}_{T-1} = \bar{d}_{T-1}, \bar{\mathbf{L}}_T = \bar{\mathbf{l}}_T) \times \\ \prod_{t=1}^{T} \; f(\mathbf{l}_t | \bar{\mathbf{A}}_{t-1} = \bar{d}_{t-1}, \bar{\mathbf{L}}_{t-1} = \bar{\mathbf{l}}_{t-1}) \end{array} \right\} d\bar{\mathbf{l}}$$

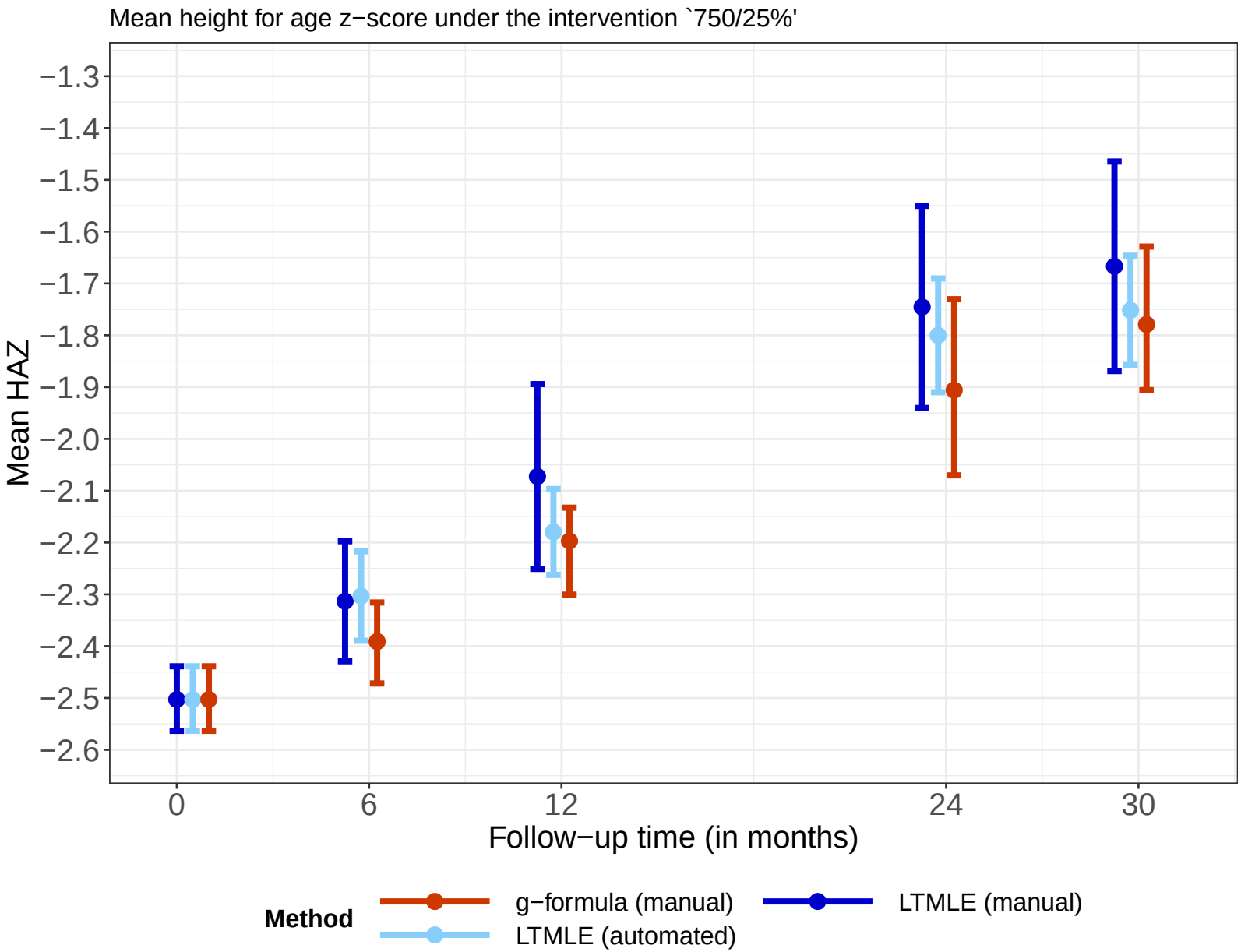
with

$$\prod_{s=1}^{q} f(l_t^s | \bar{A}_{t-1} = \bar{d}_{t-1}, \bar{\mathbf{L}}_{t-1} = \bar{\mathbf{l}}_{t-1}, L_t^1 = l_t^1, \ldots, L_t^{s-1} = l_t^{s-1}) .$$

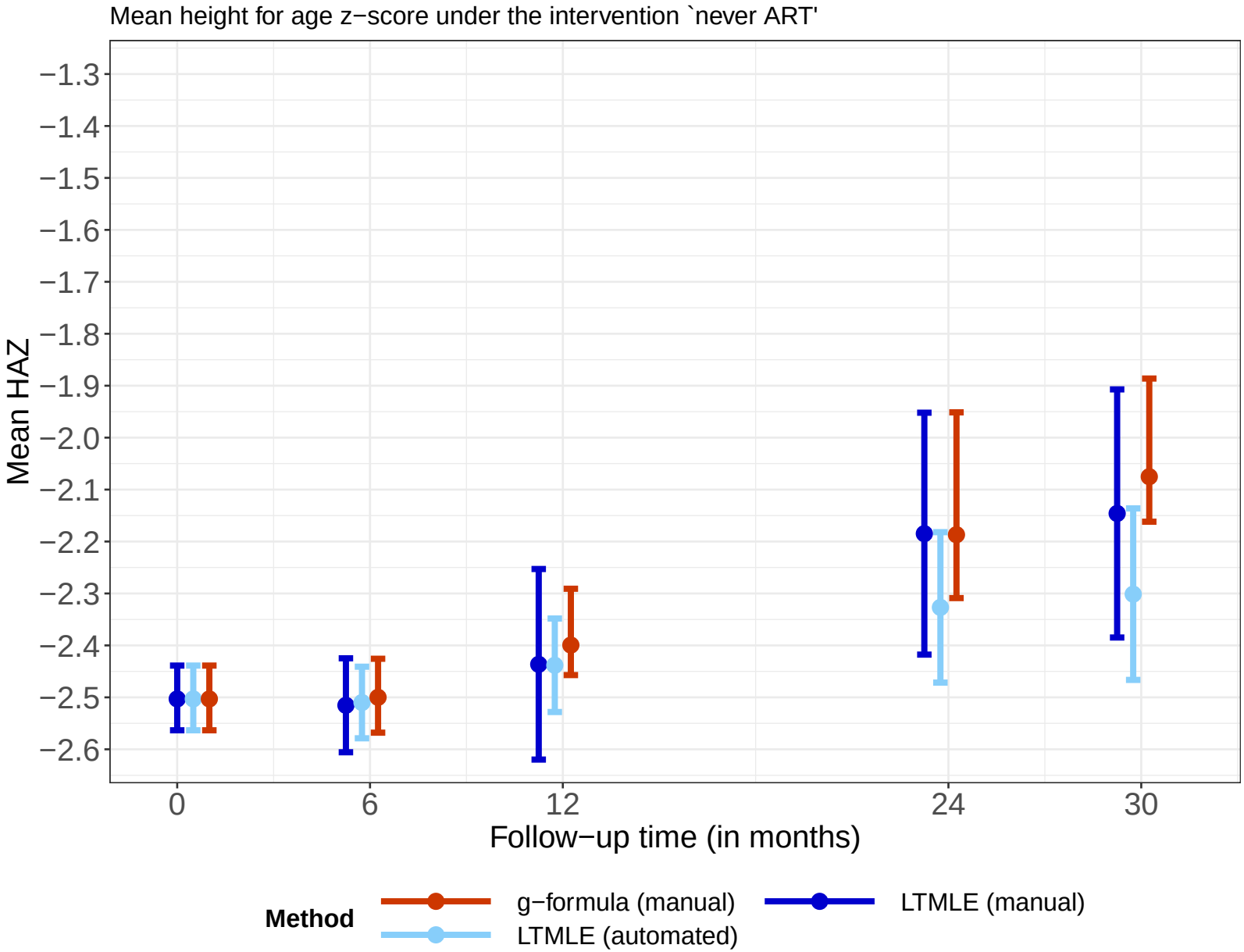
$\to$ Integral can be approximated by simulation; one requires models for all time-varying confounders L_t^s and the outcome Y_t, for $t = 1, \ldots, T$.

[3] based on Robins (1986) [4]

Results Data Analysis (Intervention 2)



Results Data Analysis (Intervention 4)

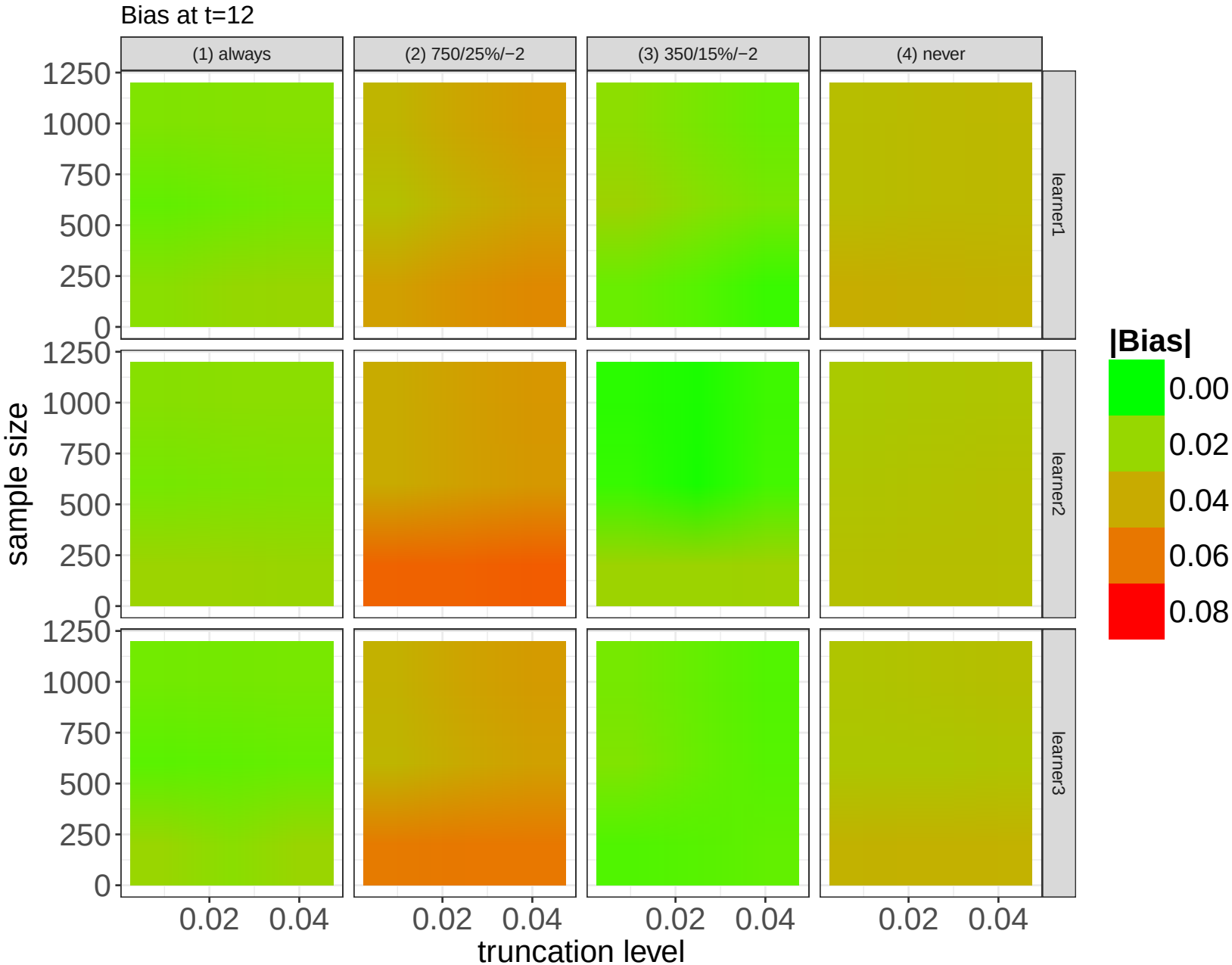


Which results should we trust?

Simulation

- There is simulation evidence that LTMLE with data-adaptive approaches performs well
- However, hardly any longitudinal settings have been evaluated; finite sample performance under multiple challenges (small sample, limited set of learners, non-linearities) unexplored
- Here: simulation with 12 time points, interactions and non-linear relationships for
  - different sample sizes; $n \in \{200, 600, 1000\}$
  - different truncation levels; $g \in \{0.01, 0.025, 0.04\}$
  - different learner sets
  - different interventions

Results Simulation



Positivity (I)

Are there problems with specific interventions? Not enough to look at crude support...

Positivity:

$$P(\mathbf{A}_t = \bar{d}_t | \bar{\mathbf{L}}_t = \bar{\mathbf{l}}_t, \bar{\mathbf{A}}_{t-1} = \bar{d}_{t-1}) > 0 \quad \textit{for} \quad \forall t, \bar{d}_t, \bar{\mathbf{l}}_t$$
$$\text{with} \quad P(\bar{\mathbf{L}}_t = \bar{l}_t, \bar{\mathbf{A}}_{t-1} = \bar{d}_{t-1}) \neq 0$$

Proposal: Estimate $P(\mathbf{A}_t = \bar{d}_t | \bar{\mathbf{L}}_t = \bar{\mathbf{l}}_t, \bar{\mathbf{A}}_{t-1} = \bar{d}_{t-1})$ to measure the *relevant* data support! (easy in simulation)

Proportion of cumulative probabilities < 0.025:

Intervention	$\bar{d}^1_{30}$	$\bar{d}^2_{30}$	$\bar{d}^3_{30}$	$\bar{d}^4_{30}$
simulation	0.3%	1.0%	0.6%	1.5%

Positivity (II) / model specification in data analysis

manual or automated LTMLE?

Proposal: calculate proportion of cumulative probabilities $< 1\%$ contained in the clever covariate for *different* model specifications.

Intervention		N	%truncated SL	GAM	GLM
$\bar{d}^1_{30}$	immediate ART	371	0	15.6	0
$\bar{d}^2_{30}$	750/25%	396	0.3	18.2	11.6
$\bar{d}^3_{30}$	350/15%	505	0	32.9	45.9
$\bar{d}^4_{30}$	no ART	292	0.7	68.5	100

$\rightarrow$ Simulation: different interventions can have different "data support"

$\rightarrow$ Diagnostics: limited data support for some interventions

$\rightarrow$ Data Analysis: caution w.r.t. interpretation of intervention 4

Conclusions

- It is feasible to implement LTMLE in complex settings with long follow-up times, small sample size, multiple time-dependent confounders, and dynamic interventions
  (first implementation with $>$ 9 follow-up time points, dynamic interventions and multiple time-dependent confounders)

- In our setting, there's no evidence that the g-formula using flexible additive models, informed by prior clinical knowledge, may perform better than an automated LTMLE procedure

- Different interventions may have different support in the data; diagnostics to detect positivity violations, as suggested, are important

Working Paper on arXiv

Schomaker M, Luque Fernandez MA, Leroy V, Davies MA.

Using Longitudinal Targeted Maximum Likelihood Estimation in Complex Settings with Dynamic Interventions.

ArXiv e-prints. 2019; `https://arxiv.org/abs/1802.05005`

Bibliography

[1] R. M. Daniel, S. N. Cousens, B. L. De Stavola, M. G. Kenward, and J. A. Sterne.
Methods for dealing with time-dependent confounding.
Statistics in Medicine, 32(9):1584–618, 2013.

[2] H. Bang and J. M. Robins.
Doubly robust estimation in missing data and causal inference models.
Biometrics, 64(2):962–972, 2005.

[3] M. J. van der Laan and S. Gruber.
Targeted minimum loss based estimation of causal effects of multiple time point interventions.
International Journal of Biostatistics, 8(1).

[4] J. Robins.
A new approach to causal inference in mortality studies with a sustained exposure period.
Mathematical Modelling, 7(9-12):1393–1512, 1986.

Appendix: DAG for Data Analysis