Comparison of Time-To-First-Event and Recurrent Event Methods in Multiple Sclerosis Trials

Alexandra Bühler, **Qing Wang**[1], **Marcel Wolbers**[1], **Fabian Model**[1], **Jan Beyersmann**[2]

[1] Roche, Biostatistics, Basel, Switzerland
[2] Ulm University, Institute of Statistics, Ulm, Germany

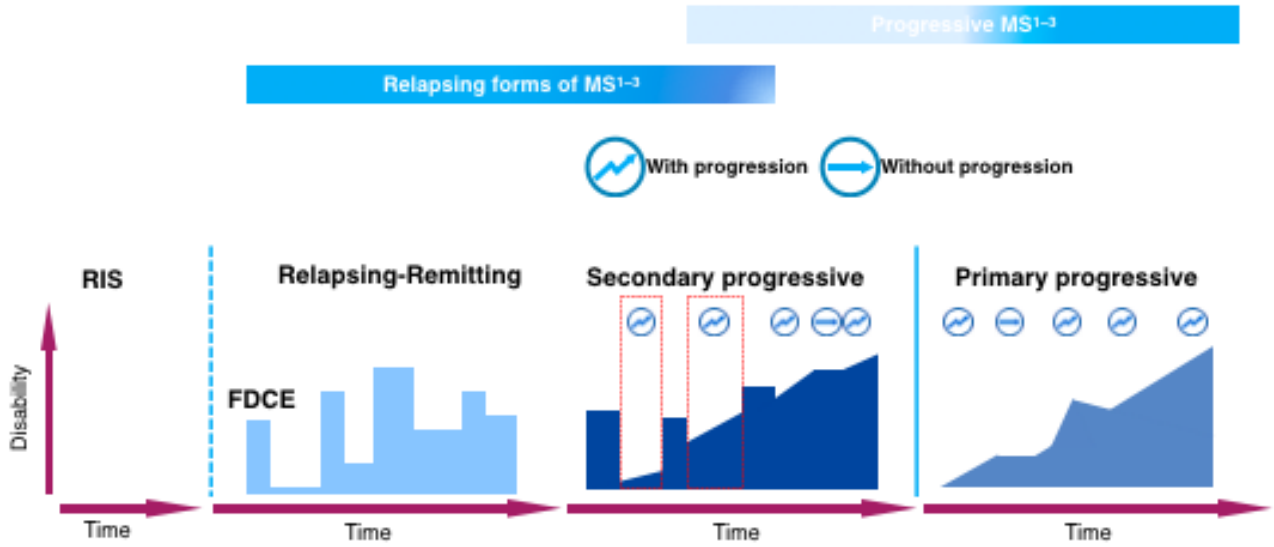
PSI Workshop, October 29, 2019

Outline

1. Introduction to MS and the ORATORIO trial in PPMS
2. Recurrent event methods in RCTs
3. Recurrent event analyses of the ORATORIO trial
4. Simulation studies
   - Generic simulation study
   - MS-specific simulation study
5. Conclusions

Multiple sclerosis (MS) disease course - 2013 consensus

- MS is a chronic, inflammatory and degenerative demyelinating disease of the human central nervous system
- Basic clinical phenomena of MS: relapses and disability progression
- Different disease courses: relapsing-remitting, secondary progressive and the primary progressive MS (RRMS, SPMS and PPMS)



FDCE, First Demyelinating Clinical Episode; RIS, radiologically isolated syndrome

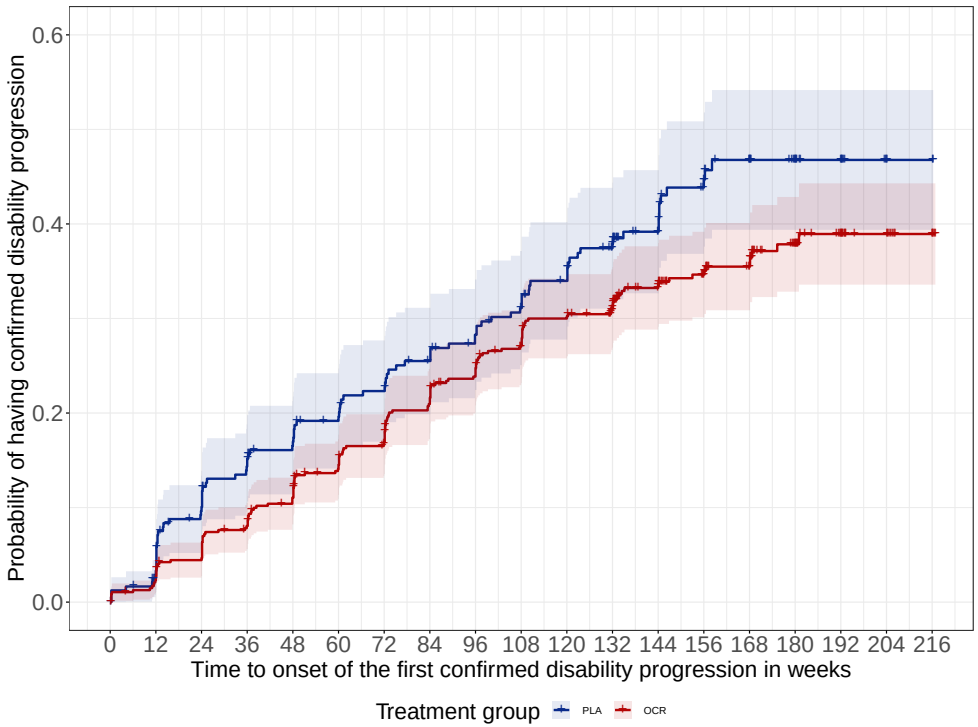
$\Longrightarrow$ Potential for recurrent event analyses in PPMS?

ORATORIO: Roche's pivotal study of ocrelizumab in PPMS

- Phase III trial
- $n = 732$ subjects, $2:1$ randomization ocrelizumab versus placebo
- **Primary endpoint:** Time from randomization to the **first** 12-week confirmed disability progression (CDP12)
  - Definition based on longitudinal assessments of Expanded Disability Status Scale (EDSS)
  - Events must be initial disability progression (IDP) which are confirmed (CDP12)
    - ★ IDP: increase in EDSS by ≥ 1.0 points (if baseline EDSS ≤ 5.5) or ≥ 0.5 points (if baseline EDSS > 5.5)
    - ★ CDP12: increase sustained for at least 12 weeks

Time-to-first-event analysis of the ORATORIO trial

Cox proportional hazards model and log-rank test:



	OCR (N=488)	PLA (N=244)
Patients included in analysis	487 (100.0 %)	244 (100.0 %)
Patients with first CDP12 event (%)	160 (32.9 %)	96 (39.3 %)
Time-to-first-event analysis		
p-value (log-rank)	0.0321	
HR (95% CI)	0.76 [0.59, 0.98]	

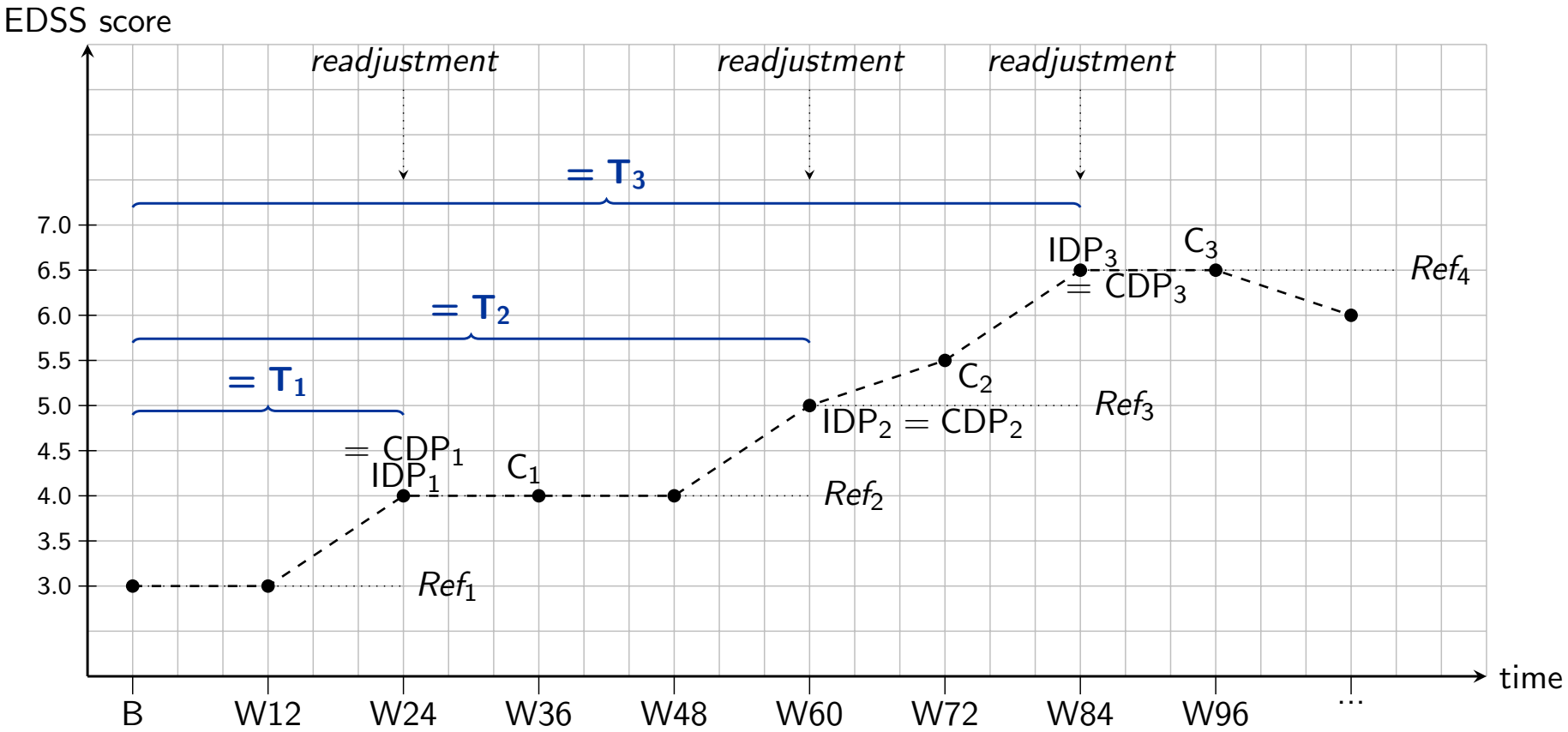
Recurrent event definition

Event	Definition
First CDP12 event	increase in EDSS score of ≥ 1.0 (if baseline EDSS ≤ 5.5) or ≥ 0.5 points (if baseline EDSS > 5.5) from the **baseline** EDSS score, confirmed for at least 12 weeks (CDP12)
Repeated CDP12 event	increase in EDSS score of ≥ 1.0 (if reference EDSS ≤ 5.5) or ≥ 0.5 points (if reference EDSS > 5.5) from a **reference** EDSS score, confirmed for at least 12 weeks (CDP12)

Reference EDSS score:

- Readjustment of reference EDSS level after each event
- Definition:
  - First CDP12 event: baseline EDSS score
  - j^{th} CDP12 event: EDSS score at IDP of $(j-1)^{th}$ CDP12 event

Recurrent event definition for a stylized subject



$IDP_j = j^{th}$ initial disability progression, C_j = confirmation of IDP_j, $CDP_j = j^{th}$ confirmed disability progression (event)
Ref_j = reference EDSS score for j^{th} CDP, T_j = time to onset of the j^{th} CDP

- Reference EDSS level readjusted after each event
- CDP definition looks into the future (as for the established time-to-first CDP definition)

Recurrent event methods

Model	Description
Negative binomial (NB)	- Parametric rate-based model - Given gamma frailty, underlying recurrent event process is Poisson - Random effect induces dependency between recurrent events - Constant event rates over time - Effect measure: RR
Lin-Wei-Yang-Ying (LWYY)	- Semiparametric rate-based model - Allows arbitrary dependence structure between recurrent events - Baseline rate function unspecified - Effect measure: RR
Andersen-Gill (AG)	- Semiparametric intensity-based model - Extension of Cox proportional hazards model - Dependence structure among recurrent events must be fully specified, (e.g., via conditioning on the past, internal time-varying covariates) - Baseline intensity function unspecified - If only adjusted for baseline covariates, underlying recurrent event process is Poisson with unspecified baseline intensity function - Effect measure: HR

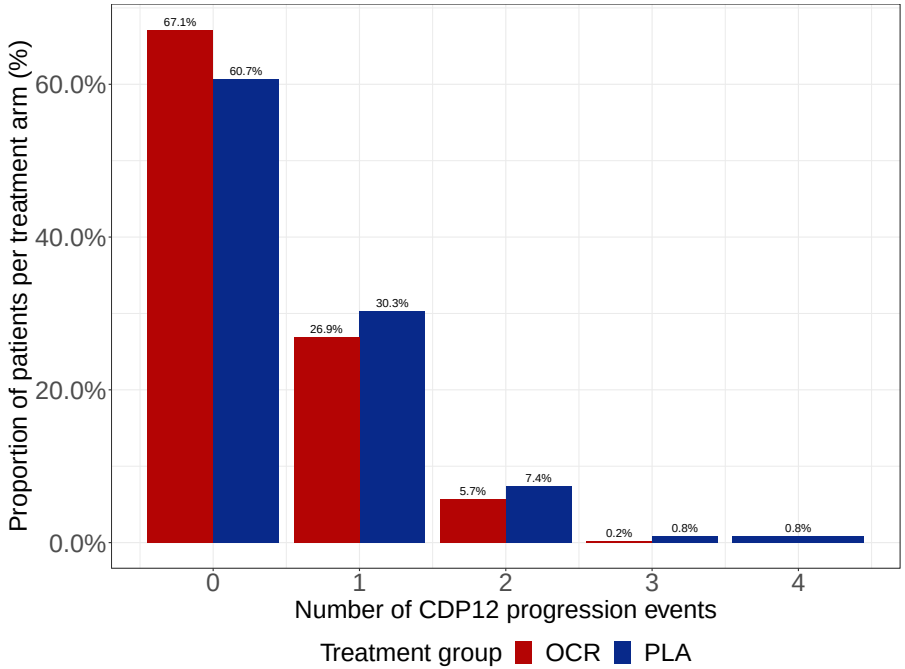
$\Longrightarrow$ All methods (NB, LWYY and AG model) estimate the 'overall' treatment effect!

Rate-based vs Intensity-based Modelling

- Conditional intensity-based model – requires full specification of the past history (event, censoring, internal/external covariate, etc.)
  - Recurrent events are conditionally uncorrelated given the past history
  - Very sensitive to model misspecification
- Marginal rate-based model – conditions only on a part of the underlying process history
  - Allows for dependence structure between recurrent events
- If the past information is incomplete, a rate function rather than a intensity function is targeted.
- The rate function can be interpreted as the average intensity function at time t across all possible histories.
- Roughly speaking: LWYY=AG with robust SE

Recurrent event analyses of the ORATORIO trial

Number of CDP12 events during double-blind treatment period:

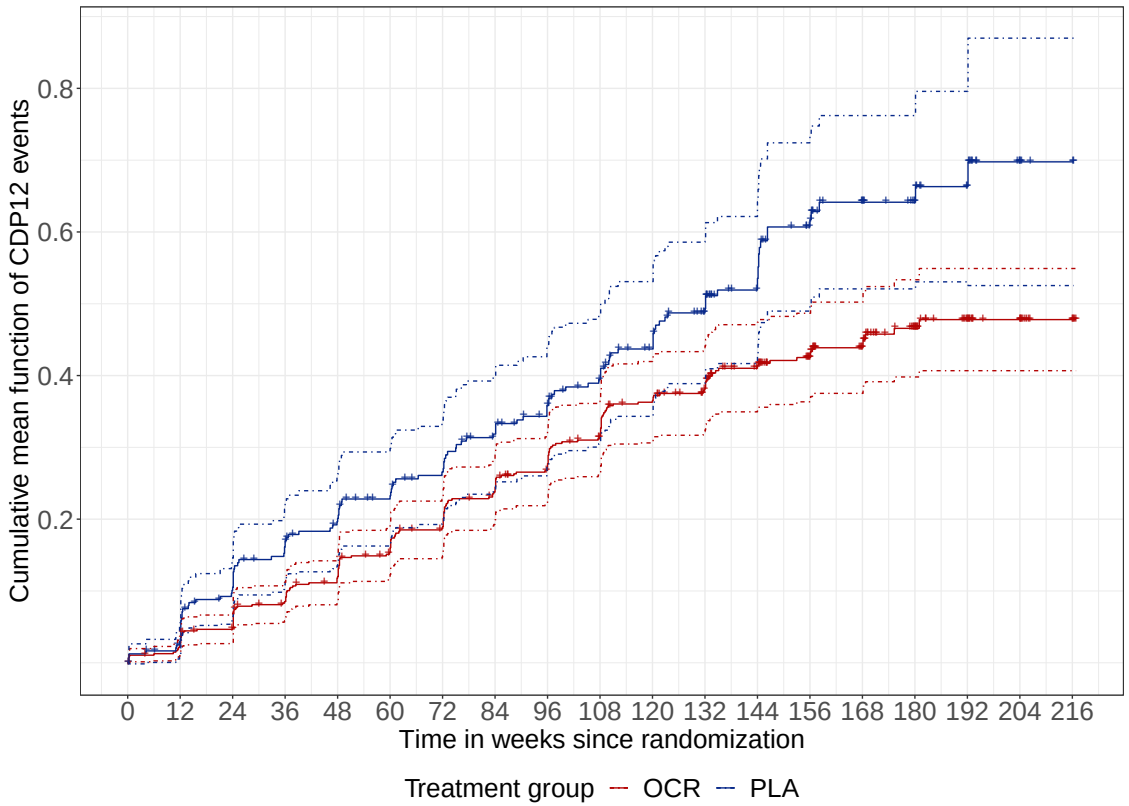


Treatment group	No. of first CDP12 events	No. of repeated CDP12 events
PLA (N=244)	96	124
OCR (N=488)	160	190
Total	256	314

$\Longrightarrow$ 58/314 (18%) CDP12 events not used in the primary time-to-first-event analysis

Recurrent event analyses of the ORATORIO trial

Cumulative mean function of CDP12 events:



- OCR or PLA patients experience on average 0.37 (95% CI $[0.31, 0.43]$) or 0.46 (95% CI $[0.36, 0.56]$) 12-week CDPs over the first 120 weeks of the double-blind treatment period

Recurrent event analyses of the ORATORIO trial

	Model	# CDP12 events included	Treatment effect
Time-to-first-event	Cox	256	HR 0.76 [0.59, 0.98], p=0.03
Recurrent event	NB	314	RR 0.71 [0.57, 0.91], p=0.005
	LWYY		RR 0.72 [0.57, 0.92], p=0.007
	AG		HR 0.72 [0.58, 0.91], p=0.005

- Smaller p-values observed for the recurrent event analyses

$\Longrightarrow$ Do all recurrent event analyses protect type I error? How much power can be gained?

$\Longrightarrow$ Investigate this in two simulation studies:

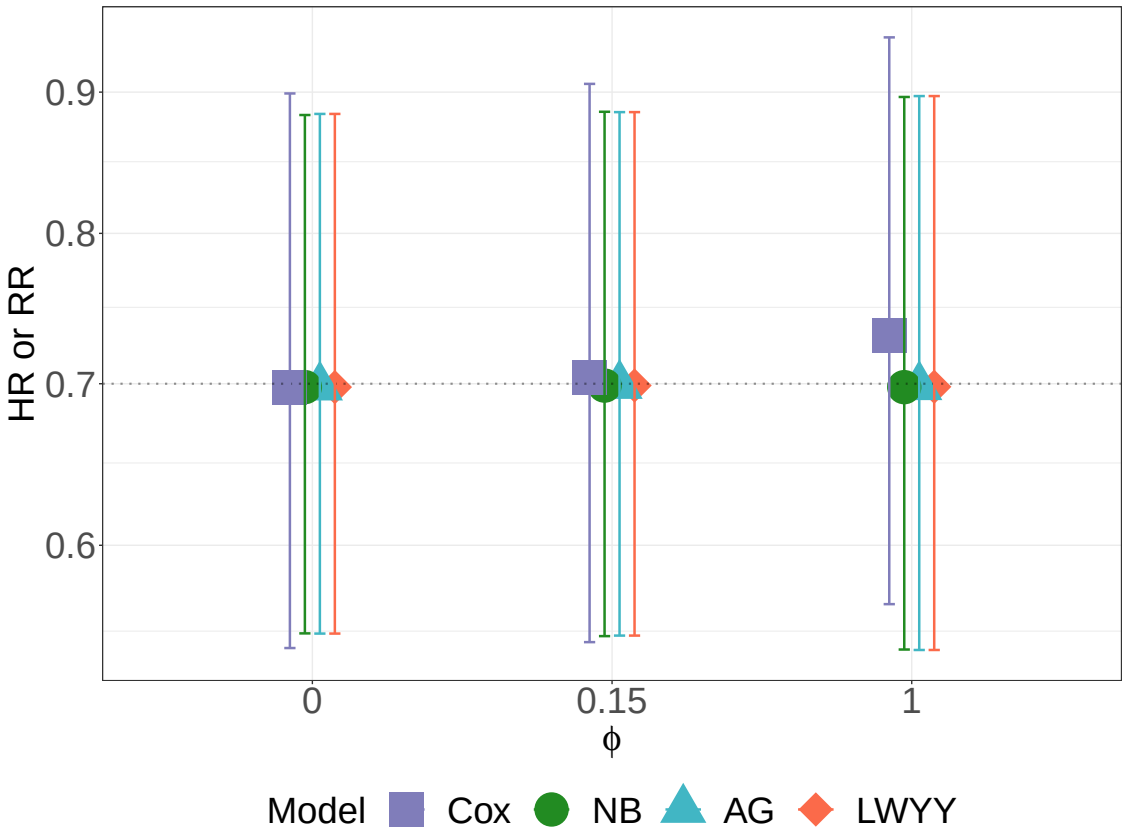
- Generic simulation study
- MS-specific simulation study

Generic simulation study - simulation setup

- Recurrent event data simulated according to a (mixed) non-homogeneous Poisson process
- Baseline intensity function of Weibull form (slightly decreasing event rates over time, as in ORATORIO data)
- Heterogeneity simulated with gamma frailty with variance ϕ:
  - homogeneous ($\phi = 0.0$), moderate ($\phi = 0.15$) and large ($\phi = 1.0$) heterogeneity
- Treatment effect: HR = 1.0 (no effect) or HR=0.7
- $n = 1000$ subjects (1:1 randomization) recruited uniformly over 1 year
- Trial continues until 246 first CDP12 events observed
  - 80% power for time to first CDP12 analysis with HR=0.7
- N = 10000 simulation runs

Generic simulation study - simulation results

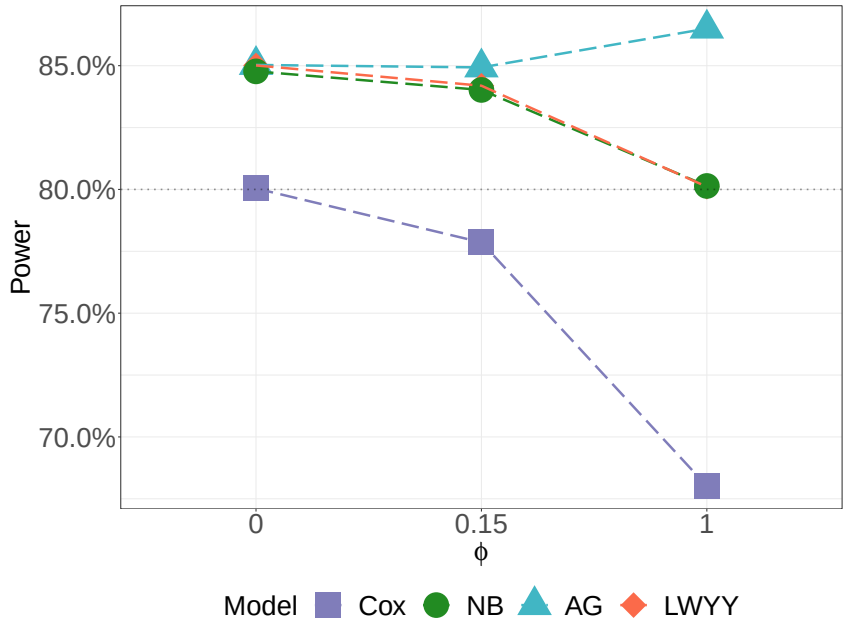
Treatment effect estimation:

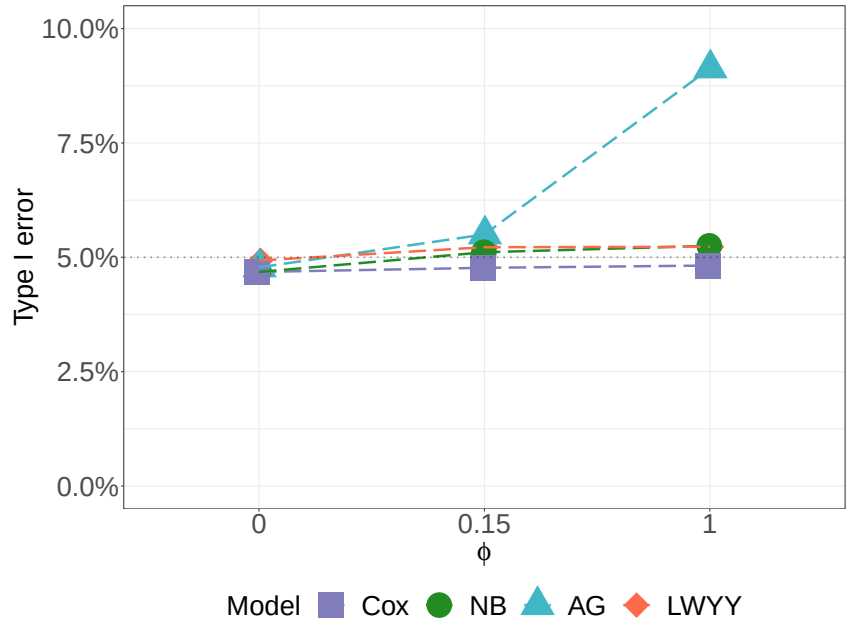


- Selection effects in time-to-first-event approach
- Recurrent event methods provide unbiased treatment effect estimates in presence of heterogeneity

Generic simulation study - simulation results

Power and type I error:



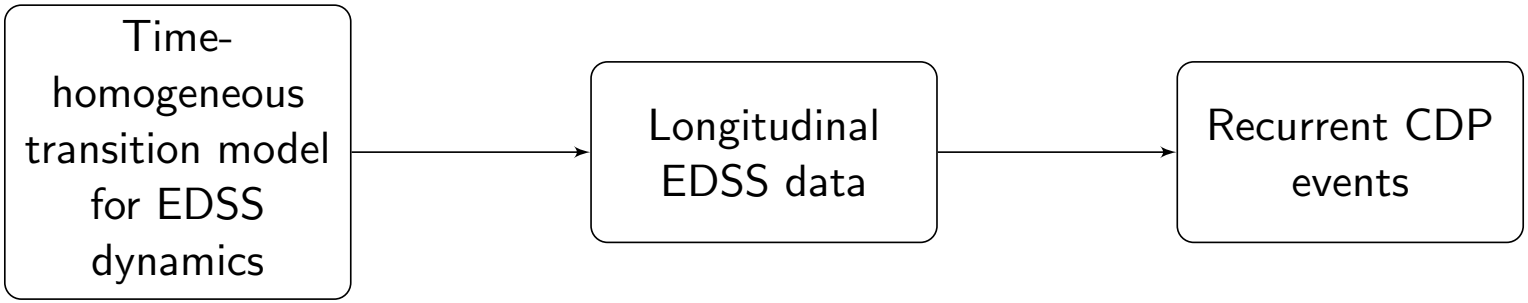


$\Longrightarrow$ Recurrent event approaches outperform time-to-first-event approach in terms of statistical power!

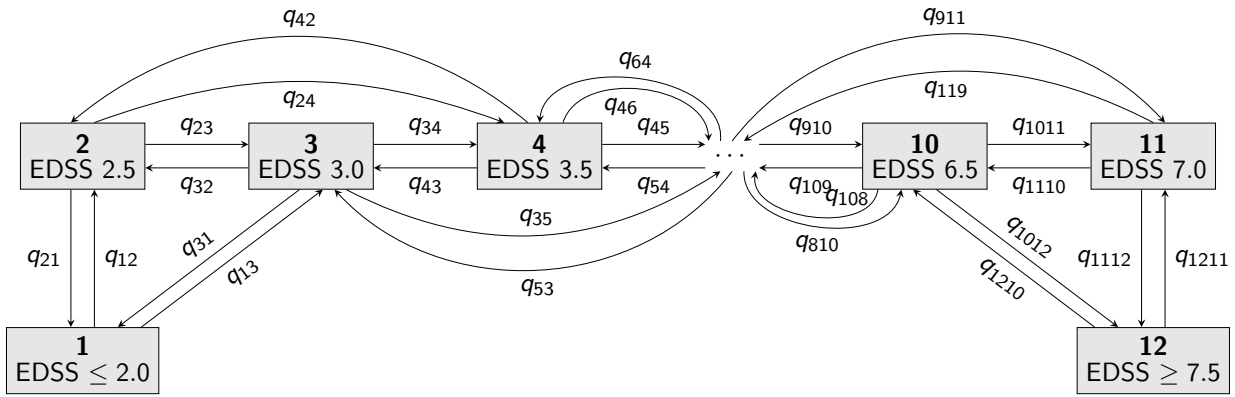
$\Longrightarrow$ Type I error inflated for the AG model in presence of heterogeneity

MS-specific simulation study - simulation setup

- Idea:



- Simulation of longitudinal ordinal EDSS measurements based on a time-homogeneous multistate model (transition intensities chosen according to ORATORIO data)

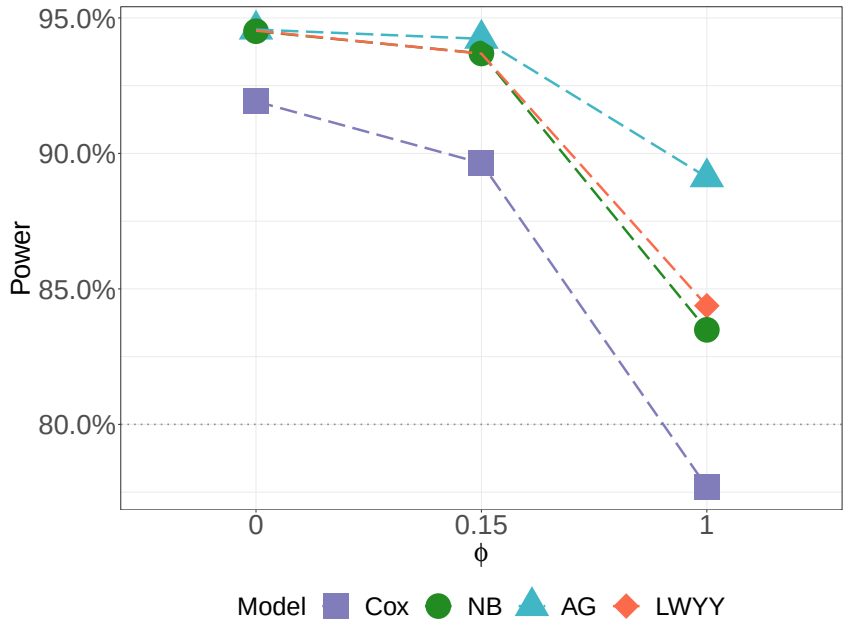


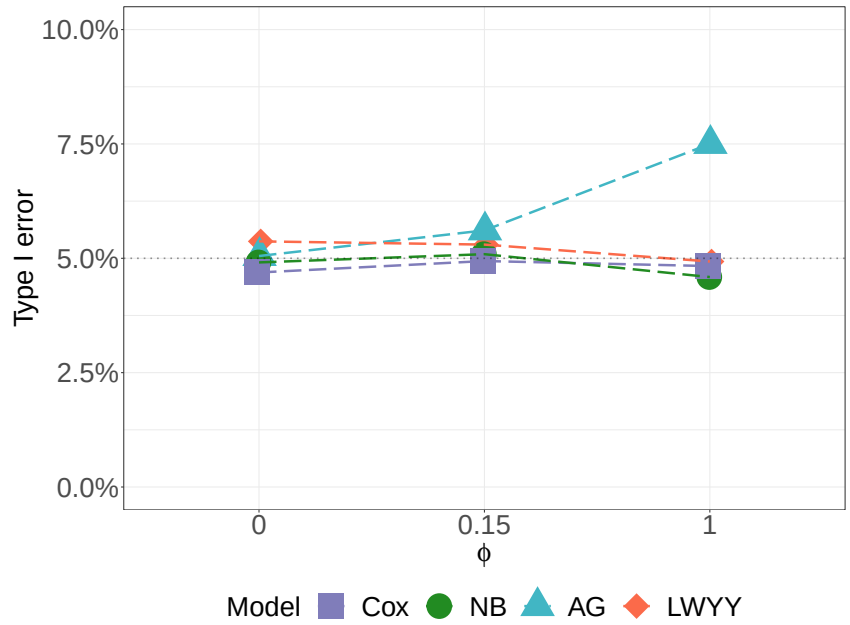
MS-specific simulation study - simulation setup

- Treatment effect and heterogeneity simulated on upper diagonal (= EDSS worsenings) of the EDSS transition intensity matrix
  - Treatment effect sizes: HR = 1.0 (no effect) or HR = 0.7
  - Gamma frailty with variance ϕ: homogeneous ($\phi = 0.0$), moderate ($\phi = 0.15$) and large ($\phi = 1.0$) heterogeneity
- Recurrent CDP endpoint derived based on simulated EDSS data
  - Simulated treatment effect sizes on transition intensity do not translate 1 : 1 to effect sizes for recurrent events
- n = 1000 subjects recruited uniformly over 1 year, trial continues until 246 first CDP events observed
- 1:1 randomization
- N = 10000 simulation runs

MS-specific simulation study - simulation results

Power and type I error:





$\Longrightarrow$ Recurrent event approaches outperform time-to-first-event approach in terms of statistical power!

$\Longrightarrow$ Type I error inflated for the AG model in presence of heterogeneity

Conclusions

- Recurrent event analysis use all clinically relevant disability progression data and increase power but add complexity
- Sample size of a trial with a recurrent endpoint could be $10 - 20\%$ lower compared to a time-to-first-event endpoint in the PPMS setting
- Type I error inflated for the AG model in the presence of heterogeneity
- Comparable performance of the LWYY and NB models
- Semiparametric LWYY model is recommened as primary analysis in RCTs
- NB model already popular for the analysis of recurrent relapses in MS
- Extension: multitype recurrent event models (CDP, 9HPT and T25FW)