

Recurrent disability progression endpoints in multiple sclerosis clinical trials

Alexandra Bühler^{ID}, Marcel Wolbers, Fabian Model, Qing Wang, Shibeshih Belachew^{ID}, Marianna Manfrini, Johannes Lorscheider, Ludwig Kappos^{ID} and Jan Beyersmann

Multiple Sclerosis Journal

2023, Vol. 29(1) 130–139

DOI: 10.1177/

13524585221125382

© The Author(s), 2022.



Article reuse guidelines:
sagepub.com/journals-
permissions

Abstract

Background: The current standard endpoint to assess disability accumulation in multiple sclerosis (MS) clinical trials is the time to the first confirmed disability progression, which excludes subsequent progression events. Including recurrent progression events may permit a more comprehensive assessment of treatment effects on disability progression.

Objective: To propose a definition of recurrent disability progression events and to compare time-to-first and recurrent event analysis.

Methods: Recurrent disability progression events were defined by expanding the recommended first event definition. Marginal recurrent event methods (negative binomial model, Lin–Wei–Yang–Ying model) were compared with Cox regression in data from three randomized controlled trials in relapsing multiple sclerosis (RMS) and primary progressive multiple sclerosis (PPMS), and in simulated randomized controlled trial data.

Results: The recurrent event analyses included a substantially larger number of progression events compared with the time-to-first-event analyses (+7.5% and +9.9% in the RMS trials and +22.7% in the PPMS trial). The increase in the number of events resulted in more precise treatment effect estimates and a corresponding gain in statistical power.

Conclusion: Our results support the use of recurrent event data analysis, especially in progressive MS trials, to improve estimates of treatment effects, increase statistical power, and better capture the clinically meaningful long-term disability progression experience.

Keywords: Ocrelizumab, multiple sclerosis, disability progression, recurrent event analyses, disease-modifying therapy, treatment outcome

Date received: 12 April 2022; revised: 26 July 2022; accepted: 18 August 2022

Introduction

Multiple sclerosis (MS) is a chronic, inflammatory, and demyelinating disease of the central nervous system that affects approximately 2.8 million people worldwide.¹ Management of MS is complex and life-long, and focuses on relapse reduction, symptom management, and prevention of long-term disability.² The most widely recognized assessment tool to quantify worsening of neurological functions and monitor changes in the level of disability over time in MS is the Expanded Disability Status Scale (EDSS), which is measured on an ordinal scale ranging from 0 to 10 in 0.5 unit increments.³ Treatment effect on disability

worsening or progression is typically determined by the time to the first disability progression event, defined as a 0.5- or 1.0-point change from baseline in EDSS score, and that needs to be confirmed by at least one additional scheduled EDSS (confirmed disability progression (CDP)).^{4,5} CDP has been used as primary or secondary endpoint in randomized clinical trials (RCTs), supporting approval of disease-modifying treatments for MS.^{6,7}

However, patients will continue to progress after reaching the first CDP event, especially those with progressive MS. The disease course after a first CDP

Correspondence to:

Q Wang

F. Hoffmann-La Roche Ltd.,
Grenzacherstrasse 124, Basel
CH-4070 Switzerland.

qing.wang.qw5@roche.com

Alexandra Bühler

University of Waterloo,
Waterloo, ON, Canada

Marcel Wolbers

Fabian Model

Qing Wang

Shibeshih Belachew

Marianna Manfrini

F. Hoffmann-La Roche Ltd.,
Basel, Switzerland

Johannes Lorscheider

Ludwig Kappos

Research Center for
Clinical Neuroimmunology
and Neuroscience Basel
(RC2NB), Departments
of Head, Spine and
Neuromedicine, Clinical
Research and Biomedical
Engineering, University
Hospital, University of Basel,
Basel, Switzerland

Jan Beyersmann

Ulm University, Ulm,
Germany

event is meaningful. For example, experiencing several CDP events in a given observation time would typically have a higher negative impact on quality of life than experiencing only one CDP event. Accounting for recurrent CDP events in estimating treatment effects promises to better capture the long-term patient experience, as well as increase statistical power.

We propose the analysis of recurrent CDP events as a comprehensive measure of disability accumulation that uses the whole within-patient trajectory of events. Recurrent events are well-established endpoints in indications with low mortality rates, including the analysis of relapses in MS.⁸ To our knowledge, analyses of recurrent CDP events have not been included as primary or secondary endpoints in MS trials to date, but they have recently been applied to observational primary progressive multiple sclerosis (PPMS) and secondary progressive MS cohorts.^{9–12}

Here, we evaluate the definition of repeated CDP events and the validity of marginal methods in measuring the benefit of an intervention on a recurrent event endpoint in an RCT setting, using data from three different RCTs in relapsing multiple sclerosis (RMS) and PPMS, as well as simulated RCT data.

Materials and methods

Definition of recurrent CDP events

We will use one of the standard definitions of EDSS-based CDP: an initial disability progression (IDP) is defined as an increase in EDSS by ≥ 1.0 point from the baseline EDSS if the baseline EDSS is ≤ 5.5 or an increase of ≥ 0.5 points if the baseline EDSS is > 5.5 .^{7,13,14}

The time to the first 12-week CDP (CDP12) is defined as the time from baseline to the onset of the first IDP that is confirmed at the next regularly scheduled visit ≥ 12 weeks thereafter. Similarly, 24-week CDP (CDP24) is defined with longer confirmation periods. Confirmation of an IDP is the gold standard methodology used to enhance specificity of the endpoint by mitigating the impact of the variability in EDSS assessments and reducing the probability of capturing progression events that may subsequently revert.

We propose to expand the definition of a first CDP event to recurrent CDP events as follows: suppose that a patient can experience up to $j = 1, 2, \dots, J$ progression events during follow-up, and let $k = 1, 2, \dots, K \geq J$ index the k th IDP since randomization, defined

as an increase in EDSS of ≥ 1.0 or ≥ 0.5 points from its reference EDSS (defined below) for reference scores ≤ 5.5 or > 5.5 , respectively. Note that $K \geq J$, since an IDP does not necessarily become a CDP. As for the first event, the k th IDP only counts as a CDP12 (or CDP24) event if it is confirmed at the next regularly scheduled visit ≥ 12 (or ≥ 24) weeks thereafter, respectively. The reference score for the k th IDP is the EDSS value associated with the IDP of the previous CDP event, and consequently must be readjusted following each identified CDP. The baseline EDSS serves as the reference score for the first event ($j = 1$), in which case the newly proposed definition of recurrent CDP events coincides with the traditional first event definition. Three examples of typical EDSS trajectories with two or three recurrent CDP12 events are shown in Figure 1. If all IDPs result in CDPs, then $K = J$ (see Figure 1(a) and (b)). Otherwise, there are fewer CDPs than IDPs (i.e. $K > J$), as in Figure 1(c).

Data sets—ORATORIO and OPERA I/II trials

Data from the two identically designed phase 3, multicenter, randomized, double-blind, double-dummy, parallel-group OPERA I and OPERA II trials (OPERA I/NCT01247324 and OPERA II/NCT01412333) in patients with RMS were used for the analyses. Study details have been reported previously.¹³ Patients ($N = 821$ and $N = 835$ in the OPERA I and II trials, respectively) were randomized (1:1) to receive either ocrelizumab (OCR) 600 mg by intravenous infusion every 24 weeks or subcutaneous interferon (IFN) β -1a three times per week at a dose of 44 μ g throughout the 96-week treatment period. The relevant institutional review boards/ethics committees approved the trial protocols, and all patients provided written informed consent.

Data from the phase 3, multicenter, randomized, double-blind, placebo (PBO)-controlled, parallel-group ORATORIO trial (NCT01194570) in patients with PPMS were also used for the analyses. Study details have been reported previously.¹⁴ Patients ($N = 732$) were randomized in a 2:1 ratio to receive OCR 600 mg ($N = 488$) or PBO ($N = 244$) every 24 weeks for at least 120 weeks until a pre-specified number of CDP12 events occurred.

Statistical analysis

A Cox regression model was used for the analysis of time to the first CDP event, with the hazard ratio (HR) as target parameter. Two marginal rate-based methods were investigated for the primary analysis of recurrent CDP events: negative binomial (NB) and

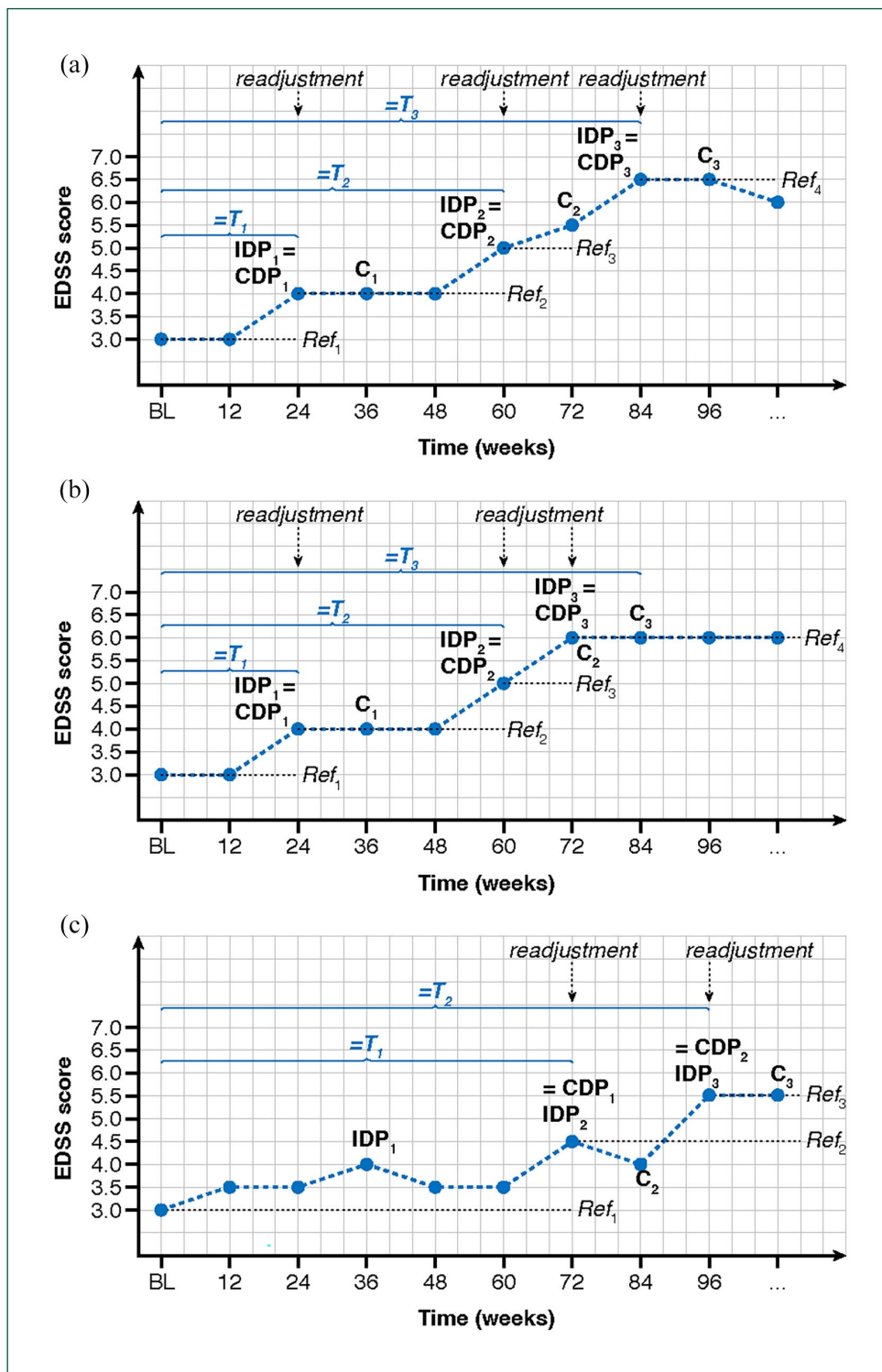


Figure 1. The definitions of first and recurrent 12-week confirmed disability progression (CDP12) events. (a) Example of three non-overlapping CDP12 events with $K=J=3$. (b) Example where confirmation of one CDP12 event (IDP_2 is confirmed at C_2) is based on the same assessment as the onset of the next CDP12 event (IDP_3 marks onset of CDP_3), with $K=J=3$. (c) Example where the first IDP_1 at week 36 cannot be confirmed and consequently does not become a CDP12 event, with $K=3$ and $J=2$.
 BL: time of randomization/baseline; CDP_j : j th 12-week confirmed disability progression for $j=1, 2, \dots, J$; C_k : confirmation of IDP_k leading to CDP_j for some $j \leq k$; EDSS: Expanded Disability Status Scale; IDP_k : k th initial disease progression for $k=1, 2, \dots, K$; Ref_j : reference disability level to which future EDSS assessments will be compared; T_j : time from baseline to j th 12-week confirmed disability progression; W: week.

Lin–Wei–Yang–Ying (LWYY) models. NB and LWYY both target the *marginal rate function*, that is, the unconditional instantaneous probability of an event at time t .¹⁵ As compared to conditional recurrent event models (e.g. Andersen–Gill, Prentice–Williams–Peterson), they retain the beneficial aspects of randomization, provide treatment effect estimates with a causal interpretation, and are therefore recommended for the primary analysis of an RCT with recurrent CDP events.^{16,17}

The NB method models a subject's number of CDP events and includes their (log-transformed) follow-up duration as an offset. It assumes that the rate function is constant over time and, if this is fulfilled, then the resulting treatment effect estimate is a simple multiplicative factor and can be interpreted as a rate ratio (RR). Unobserved heterogeneity between subjects is modeled via gamma-distributed random effects, often referred to as frailties.¹⁸ Sample size formulas for fixed and group-sequential designs with NB outcomes have been published.¹⁹

The LWYY method is one of the most commonly used models for recurrent event analysis in RCTs with negligible mortality.¹⁵ It allows for a varying rate function over time and assumes that the treatment acts multiplicatively on the rate function. No assumptions regarding the baseline rate function or the dependence structure between the recurrent events are required, but it typically assumes recurrent event process and time to censoring to be independent (given covariates). The target parameter is a ratio of rate functions across the treatment groups (i.e. RR), and inference is based on a robust sandwich variance estimator. For a comprehensive discussion of these methods, we refer to the Supplement, Cook and Lawless, and Bühler.^{16,17}

Design of the simulation study

A simulation study was set up to compare the power and type I error rate of the time-to-first and recurrent event methods in a controlled environment. The choice of simulation parameters (including event generating process) was based on the ORATORIO trial in which all patients were followed until a pre-determined number of first CDP12 events had occurred, resulting in a median follow-up of 29.8 months. We simulated hypothetical 1:1 randomized trials, which enrolled 1,000 subjects with PPMS uniformly over 12 months and followed them until 246 first CDP events were observed. This number of events guarantees a power of 80% for the time-to-first-event Cox analysis assuming a marginal HR of 0.7. CDP events were simulated

according to a (mixed) non-homogeneous Poisson process, where the baseline function was of Weibull shape and slightly decreasing over time. A subject's frailty term was generated from a gamma distribution whose variance was varied between 0 (no heterogeneity), 0.15 (mild heterogeneity corresponding to the impact of baseline prognostic factors observed in ORATORIO), and 1 (large heterogeneity). The subject's treatment effect was chosen as either 1 (no effect) or 0.7 (beneficial effect) on the conditional HR scale (see also the Supplemental Material).

Patient drop-out was modeled using an exponential distribution with an annual drop-out probability of 8.7%. For each of the simulation scenarios, the proportion of rejected null hypotheses among 10,000 simulation runs was summarized for all analysis methods. This proportion corresponds to the type I error for a treatment effect of 1 and to the power for the treatment effect of 0.7.

Full details of the simulation study are reported in Bühler.¹⁷ All simulations and other analyses were conducted with the statistical software R, version 3.5.2.

Results

OPERA I/II and ORATORIO trials

We have applied the above definition for recurrent CDP events to the OPERA and ORATORIO trials. In the OPERA I trial, 50 and 31 patients in the IFN and OCR treatment arms, respectively, experienced at least one CDP12 event, and a total of 56 and 33 CDP12 events, corresponding to an increase in events of 12.0% and 6.5%, respectively (Table 1). In the OPERA II trial, 63 and 44 patients in the IFN and OCR treatment arms, respectively, experienced at least one CDP12 event, and a total of 69 (+9.5%) and 46 (+4.5%) CDP12 events. In the ORATORIO trial, 96 and 160 patients in the PBO and OCR treatment arms, respectively, experienced at least one CDP12 event, and a total of 124 (+29.2%) and 190 (+18.8%) CDP12 events (Table 1). Similarly increased yield of events was observed for CDP24 (Table 1).

For time-to-first-event analyses, the accrual of events over time is typically visualized using the Kaplan–Meier methods, where the y -axis shows the estimated proportion of patients with a first event out of the overall study population (Figure 2(a), (c), and (e)). For recurrent event analyses, the accrual of events over time is typically shown as cumulative mean function, where the y -axis shows the estimated

Table 1. Distribution of first and recurrent 12- and 24-week confirmed disability progression events in the OPERA and ORATORIO trials.

Study (population)	Treatment group	Number of CDP12 events, <i>n</i>					First events, <i>n</i>	Total events, <i>n</i>	Increase, %
		0	1	2	3	4			
OPERA I (RMS)	IFN (<i>N</i> =411)	360	44	6	0	0	50	56	12.0
	OCR (<i>N</i> =410)	379	30	0	1	0	31	33	6.5
OPERA II (RMS)	IFN (<i>N</i> =418)	355	58	4	1	0	63	69	9.5
	OCR (<i>N</i> =417)	373	42	2	0	0	44	46	4.5
ORATORIO (PPMS) ^a	PBO (<i>N</i> =244)	148	74	18	2	2	96	124	29.2
	OCR (<i>N</i> =488)	327	131	28	1	0	160	190	18.8

Study (population)	Treatment group	Number of CDP24 events, <i>n</i>					First events, <i>n</i>	Total events, <i>n</i>	Increase, %
		0	1	2	3	4			
OPERA I (RMS)	IFN (<i>N</i> =411)	371	33	6	0	0	39	45	15.4
	OCR (<i>N</i> =410)	386	23	1	0	0	24	25	4.2
OPERA II (RMS)	IFN (<i>N</i> =418)	370	45	2	1	0	48	52	8.3
	OCR (<i>N</i> =417)	384	32	1	0	0	33	34	3.0
ORATORIO (PPMS) ^a	PBO (<i>N</i> =244)	157	67	16	3	1	87	112	28.7
	OCR (<i>N</i> =488)	343	116	27	1	0	144	173	20.1

CDP12/24: 12/24-week confirmed disability progression; EDSS: Expanded Disability Status Scale; IFN: interferon; OCR: ocrelizumab; PBO: placebo; PPMS: primary progressive multiple sclerosis; RMS: relapsing multiple sclerosis; increase, %: total events/first events × 100.
^aPatient with missing baseline EDSS excluded from analysis.

average number of events per patient in the overall study population (Figure 2(b), (d), and (f)). By Week 96 of the OPERA I study, 13.0% and 8.3% of patients in the IFN and OCR arm, respectively, had experienced at least one CDP12 event, and the average number of CDP12 events per patient was 0.14 and 0.09, respectively. By Week 96 of the OPERA II study, 17.5% and 11.1% of patients in the IFN and OCR arm, respectively, had experienced at least one CDP12 event, and the average number of CDP12 events per patient was 0.20 and 0.12, respectively. By Week 120 of the ORATORIO study, 34.0% and 30.2% of patients in the PBO and OCR arm, respectively, had experienced at least one observed CDP12 event, and the average number of observed CDP12 events per patient was 0.46 and 0.37, respectively. Event rates in all three trials have been found to be approximately constant, justifying the use of the NB model.

The results using the different models are summarized for all three trials in Table 2. For all trials and confirmation period durations (12 and 24 weeks), the treatment effect on the recurrent events was comparable to the time-to-first-event analysis, the recurrent event analyses being associated with narrower confidence intervals (CIs).

For the ORATORIO trial, we additionally explored the impact of treatment on a subject's first, second, or

third CDP event using partially conditional rate-based (PCRB) models.^{16,17} The PCRB model with a common treatment effect is similar to the LWYY model, but additionally adjusts for event-specific baseline rate functions. The event-specific PCRB analysis for the first, second, or third CDP event, respectively, gives an estimate of the treatment effect among all subjects (for the first event) or among those subjects who have already experienced one or two earlier CDP events. For both confirmation period durations (12 and 24 weeks), the observed treatment effects for secondary and later events are in close agreement with the time-to-first-event results (Table 2), and so the treatment effect was found to be consistent across the first, second, and later CDP events. However, by conditioning on the occurrence of previous CDP events, the PCRB model no longer compares randomized treatment groups, and hence, the resulting treatment effect estimates lack a formal causal interpretation. Such an exploratory analysis is nevertheless meaningful in providing a supplementary characterization of the treatment effect.

Simulation study

For the simulation scenarios with a treatment effect of 0.7, the median study duration, that is, the time until the targeted 246 first CDP events had been observed, was 27, 28, and 32 months, respectively, assuming no,

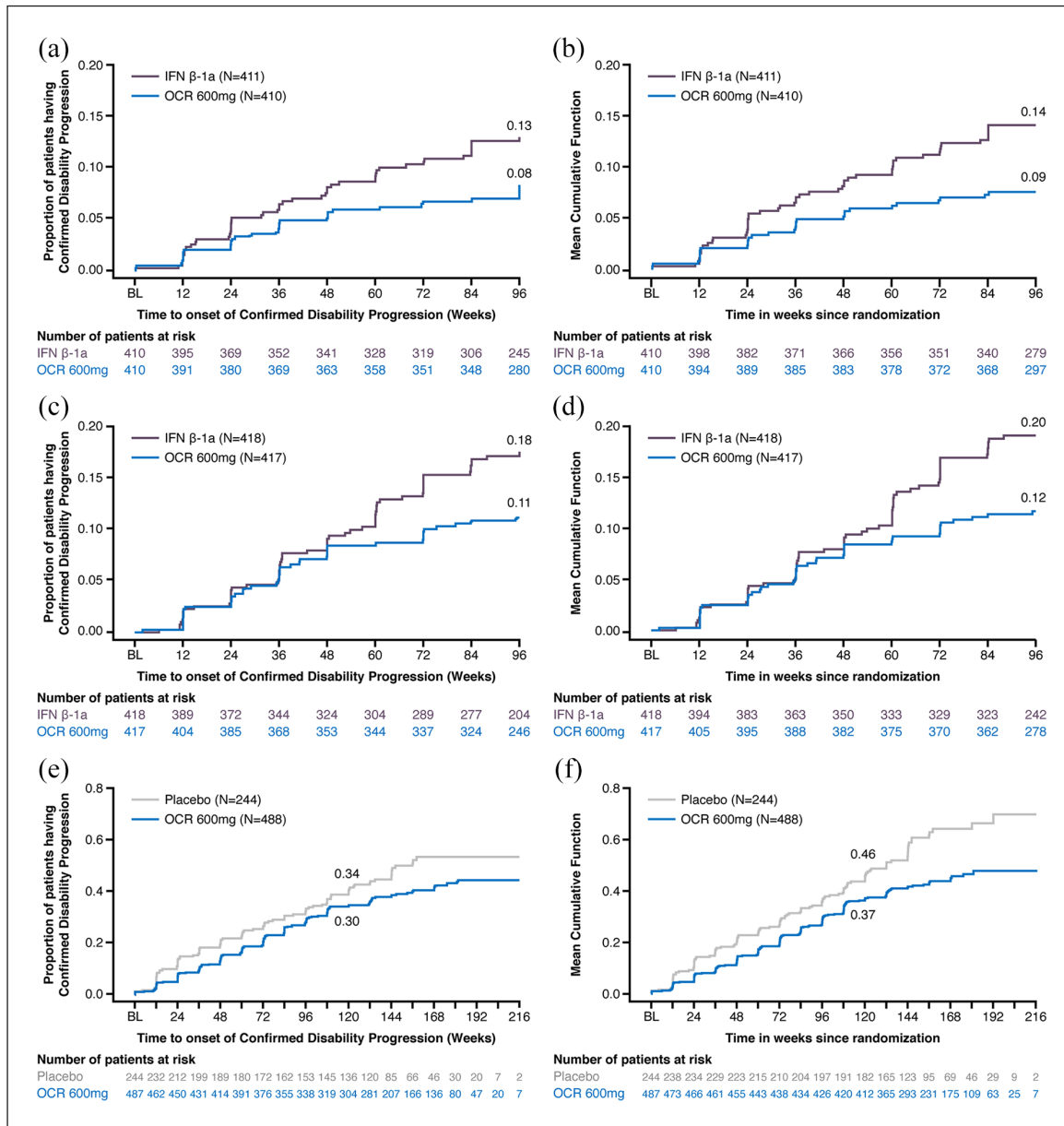


Figure 2. Cumulative risk of a first 12-week confirmed disability progression (1 – Kaplan–Meier estimate; left panels) and mean cumulative number of 12-week confirmed recurrent disability progressions (Nelson–Aalen type estimate; right panels) over time in the (a, b) OPERA I; (c, d) OPERA II; and (e, f) ORATORIO trials. BL: time of randomization/baseline IFN: interferon; OCR: ocrelizumab.

mild, or large heterogeneity. The corresponding median number of all CDP events (compared with the number of first events) was 286 (+16%), 293 (+19%), and 335 (+36%), respectively.

The average treatment effect for the time-to-first CDP was 0.700, 0.704, and 0.730 for no, mild, or large heterogeneity, respectively. In contrast, all recurrent event methods gave an average treatment effect of 0.70, that is, they were approximately unbiased for

the “real” RR=0.7. If there was no heterogeneity, LWYY, NB, and Cox analyses result in estimates that are consistent for the same parameter (i.e. RR=HR=0.7). However, if there is between-subject heterogeneity, parameters from these analyses differ, and treatment effects for the time-to-first-event Cox analysis are increasingly attenuated with increasing heterogeneity. This can be explained by a selection effect: as the follow-up time increases, an increasing number of subjects prone to worsening are withdrawn

Table 2. Treatment effect estimates of different first and recurrent event models for 12- and 24-week confirmed disability progression events in the OPERA and ORATORIO trials.

Study (population)	Model	12-week CDP		24-week CDP	
		Treatment effect HR/RR (95% CI)	<i>p</i> -value	Treatment effect HR/RR (95% CI)	<i>p</i> -value
OPERA I ^a	Cox	0.574 (0.366–0.899)	0.0153	0.569 (0.342–0.947)	0.0301
	LWYY	0.566 (0.361–0.888)	0.0132	0.530 (0.319–0.881)	0.0143
	NB	0.558 (0.356–0.874)	0.0108	0.528 (0.314–0.886)	0.0156
OPERA II ^a	Cox	0.626 (0.425–0.923)	0.0182	0.627 (0.402–0.977)	0.0391
	LWYY	0.609 (0.419–0.886)	0.0094	0.608 (0.395–0.937)	0.0241
	NB	0.615 (0.424–0.894)	0.0108	0.604 (0.391–0.935)	0.0236
ORATORIO ^b	Cox	0.759 (0.589–0.978)	0.0330	0.754 (0.577–0.984)	0.0375
	LWYY	0.723 (0.572–0.915)	0.0070	0.731 (0.569–0.940)	0.0145
	NB	0.714 (0.564–0.904)	0.0051	0.717 (0.554–0.928)	0.0115
	PCRB				
	Common	0.752 (0.599–0.946)	0.0147	0.767 (0.602–0.977)	0.0315
	Event 1	0.759 (0.590–0.977)	0.0322	0.754 (0.578–0.983)	0.0372
	Event 2	0.776 (0.445–1.356)	0.3738	0.911 (0.507–1.635)	0.7542
	Event ≥3	0.156 (0.019–1.307)	0.0867	0.156 (0.019–1.307)	0.0867

CDP: confirmed disability progression; CI: confidence interval; EDSS: Expanded Disability Status Scale; HR: hazard ratio; LWYY: Lin–Wei–Yang–Ying; NB: negative binomial; PCRB: partially conditional rate-based; RR: rate ratio.

^aHR (Cox) and RR (LWYY) were estimated by stratified models, with different baseline hazard or rate functions for each strata defined by baseline Expanded Disability Status Scale (EDSS) category (<4 vs ≥4) and geographical region (the United States vs rest of the world). RR under NB model adjusted for these randomization stratification factors.

^bHR (Cox) and RR (LWYY and PCRB) were estimated by stratified models, with different baseline hazard or rate functions for each strata defined by baseline age group (>45 vs ≤45 years) and geographical region (the United States vs rest of the world). RR under NB model adjusted for these randomization stratification factors.

from the risk set, and only the subjects less prone to worsening remain. This selection does not occur for recurrent event analyses, as follow-up of subjects after their first event is also included in the analysis. For the simulation scenarios with no treatment effect, all methods were approximately unbiased for HR=1.0 or RR=1.0, assuming no, mild, or large heterogeneity.

The estimated power and type I error rate for all simulation scenarios considered are shown in Figure 3(a) and (b). The power of the Cox model is 80% assuming no heterogeneity but deteriorates to 68% for large heterogeneity. The power of the NB and LWYY models is similar and ranges from 85% (no heterogeneity) to 80% (large heterogeneity). All models maintain the type I error rate at approximately 5%.

In simulations without heterogeneity, the power of the recurrent event analysis compared with the time-to-first-event analysis increased from 80% to 85%. A recurrent event analysis with 80% power would require approximately 13% less recurrent events, and thus 13% less subjects than a time-to-first-event analysis for the same power and study duration. For scenarios with a higher heterogeneity, the power gains

associated with a recurrent event analysis were larger. As a consequence, 13% is a conservative estimate of the sample size savings associated with recurrent event analysis in our scenarios.

Discussion

Clinical trials in MS entail large sample sizes and long follow-up periods in order to generate reliable results with regards to the effect of an intervention.⁴ Disability progression measured by the EDSS is a key efficacy outcome used in phase 3 MS clinical trials.⁵ Given that disability will continue to worsen after the occurrence of the first CDP event,²⁰ especially in progressive forms of MS, we demonstrated that the analysis of recurrent CDP events after the first event increases the statistical power of a study, and captures clinically meaningful events more comprehensively.

In the recurrent event analyses, the number of CDP12 events in the two RMS studies increased by 7.5%–9.9% compared with the respective time-to-first-event analyses; the corresponding relative increase in the PPMS study was 22.7%. The simulation study, that mimicked a PPMS trial, confirmed that the recurrent event analysis confers gains in statistical power. We

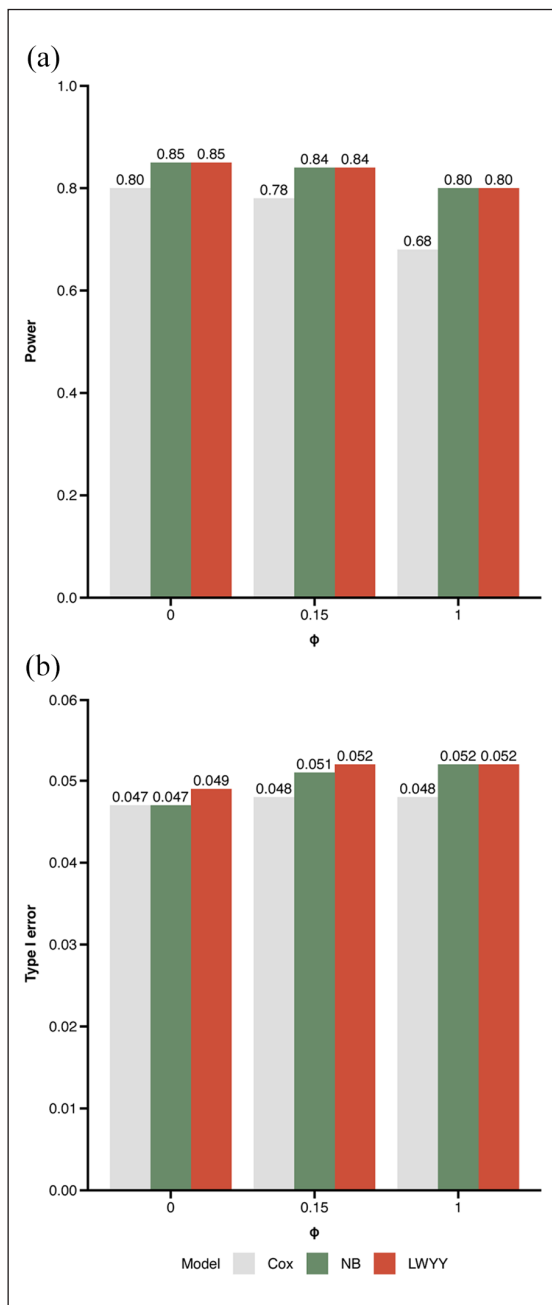


Figure 3. (a) Power and (b) type I error for all simulation scenarios depending on the between-subject heterogeneity parameter ϕ . The intermediate value of $\phi=0.15$ corresponds to the heterogeneity observed in the ORATORIO trial. Results are based on 10,000 simulations per scenario implying an approximate precision (standard error) of ± 0.004 for the simulated power and of ± 0.002 for the simulated type I error, respectively. LWYY: Lin–Wei–Yang–Ying; NB: negative binomial.

conservatively estimated that a trial powered to a recurrent event endpoint would require at least 13% less subjects compared with a trial powered for the time-to-first-event analysis. In contrast, the benefits

of recurrent event analyses would be less apparent in RMS trials with fewer recurrent events observed. Nevertheless, the recurrent event methodology might have meaningful advantages in RMS trials with longer follow-up.

It is important that appropriate statistical methods are used to compare recurrent events in randomized trials. Existing recurrent event analysis methods differ in that they either model the conditional hazard function (also known as the intensity function), that is, a subject's instantaneous probability of a new event at any time conditional on the entire process history, or the marginal rate function. The popular, intensity-based Andersen–Gill and Prentice–Williams–Peterson models are powerful frameworks for the descriptive analysis of complex recurrent event processes, but do not support causal inferences in randomized trials.^{16,17} For the primary analysis of an RCT, marginal rate-based models (NB and LWYY) are recommended because they retain the randomized treatment group analysis and hence allow for causal effect interpretation.

In our simulation study, we showed that both methods (NB and LWYY) controlled the type I error rate and had comparable power. Currently, the NB model is the recommended method for the analysis of annualized relapse rates in MS trials.¹⁸ An advantage of the LWYY model is that it is valid regardless of the shape of the baseline rate function, unlike the NB model, which assumes a constant rate function. In practice, since event rates in MS patients were found to be approximately constant over time at least in the studies analyzed here, both models can be chosen for the primary analysis, with the other as a sensitivity analysis.

The use of the proposed recurrent event methodology requires that all patients stay on randomized treatment for the full duration of the trial. In situations where effective rescue treatments for progressing patients are available, it might be ethically questionable for patients to stay on a potentially ineffective therapy after the first CDP. Hence, an inherent limitation of the proposed recurrent event method is that it will be more affected by premature treatment discontinuations and study withdrawals than a time-to-first-event analysis. One possible approach to mitigate this disadvantage is to ignore later rescue medication and analyze patients according to the intention-to-treat intention-to-treat principle. Another approach would be to analyze premature treatment discontinuations and study withdrawals as a competing risk using recurrent event methodology accounting for competing risks.^{21,22}

We extended the standard definitions of EDSS-based CDP to recurrent events. For individual subjects, EDSS trajectories over time may be quite variable and show alternation of improvement and worsening events complicating their interpretation. Other MS disability endpoints including objective performance outcome measures, for example, the time to the first confirmed progression in functional impairment of ambulation and manual dexterity, as defined by changes in the Timed 25-Foot Walk (T25FW) or the 9-Hole Peg Test (9HPT),^{23,24} respectively, or in cognitive tests like the Symbol Digit Modalities Test (SDMT), are also frequently reported in clinical studies, and may be characterized by less random variation and measurement error in particular in progressive MS. Future extension of the recurrent event analysis methodology to include additional and ideally more objective MS performance outcome measures should be considered.

Recurrent event methods enable the use of patient outcome data after the occurrence of first CDP for statistical inference and have wide applications. Beyond RCTs, recurrent event methods are also promising for both long-term extension studies and observational cohort analyses with long follow-up periods, where after a few years of follow-up, most patients will have experienced a first CDP. In these studies, the long-term quantification of the cumulative mean or the annual rate of progression using recurrent event analysis might enable more comprehensive capture of disability experience in patients' disease trajectories.

Acknowledgements

The authors thank all patients, their families, and the investigators who participated in the trials contained within these analyses. Writing and editorial assistance was provided by Martha Hoque of Articulate Science, UK, and funded by F. Hoffmann-La Roche Ltd.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: A.B. completed an internship at F. Hoffmann-La Roche Ltd. while researching the submitted work. During completion of the work related to this manuscript, A.B. was an MSc student at the Ulm University, 89081 Ulm, Germany. M.W. is an employee and shareholder of F. Hoffmann-La Roche Ltd. F.M. is an employee and shareholder of Denali Therapeutics. During completion of the work related to this manuscript, F.M. was an employee of F. Hoffmann-La Roche Ltd. Q.W. is an employee of F. Hoffmann-La Roche Ltd. S.B. is an employee and shareholder of Biogen Inc. During

completion of the work related to this manuscript, S.B. was an employee of F. Hoffmann-La Roche Ltd. M.M. is an employee and shareholder of F. Hoffmann-La Roche Ltd. J.L. has received research support from Innosuisse—Swiss Innovation Agency, Biogen, and Novartis; and speaking honoraria and/or fees for serving on advisory boards from Novartis, Roche, and Teva. L.K.s' institution (University Hospital Basel) has received the following exclusively for research support: steering committee, advisory board, and consultancy fees from: Actelion, Bayer HealthCare, Biogen, BMS, Genzyme, GlaxoSmithKline, Janssen, Japan Tobacco, Merck, Novartis, Roche, Sanofi, Santhera, Shionogi, TG Therapeutics; speaker fees from: Bayer HealthCare, Biogen, Merck, Novartis, Roche, and Sanofi; support of educational activities from: Allergan, Bayer HealthCare, Biogen, CSL Behring, Desitin, Genzyme, Merck, Novartis, Roche, Pfizer, Sanofi, Shire, and Teva; license fees for Neurostatus products and grants from: Bayer HealthCare, Biogen, European Union, InnoSwiss, Merck, Novartis, Roche, Swiss MS Society, and Swiss National Research Foundation. J.B.'s institution has received a research collaboration grant from F. Hoffmann-La Roche Ltd. not related to the current research.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by F. Hoffmann-La Roche Ltd., Basel, Switzerland.

ORCID iDs

Alexandra Bühler ID  <https://orcid.org/0000-0002-0872-0374>

Shibeshih Belachew ID  <https://orcid.org/0000-0003-3976-1950>

Ludwig Kappos ID  <https://orcid.org/0000-0003-4175-5509>

Data Sharing Statement

For eligible studies, qualified researchers may request access to individual patient-level clinical data through a data request platform. At the time of writing this, request platform is Vivli (<https://vivli.org/ourmember/roche/>). For up to date details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here https://go.roche.com/data_sharing. Anonymised records for individual patients across more than one data source external to Roche cannot, and should not, be linked due to a potential increase in risk of patient re-identification.

Supplemental Material

Supplemental material for this article is available online.

References

- Multiple Sclerosis International Federation. *Atlas of multiple sclerosis* (3rd ed.), <https://www.atlasofms.org/map/global/epidemiology/number-of-people-with-ms> (2020, accessed 8 February 2021).
- Murray TJ. Diagnosis and treatment of multiple sclerosis. *BMJ* 2006; 332: 525–527.
- Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An Expanded Disability Status Scale (EDSS). *Neurology* 1983; 33(11): 1444–1452.
- Ebers GC, Heigenhauser L, Daumer M, et al. Disability as an outcome in MS clinical trials. *Neurology* 2008; 71: 624–631.
- BMJ Uitdehaag. Disability outcome measures in phase III clinical trials in multiple sclerosis. *CNS Drugs* 2018; 32: 543–558.
- van Munster CE and Uitdehaag BM. Outcome measures in clinical trials for multiple sclerosis. *CNS Drugs* 2017; 31: 217–236.
- European Medicines Agency. Guideline on clinical investigation of medicinal products for the treatment of multiple sclerosis, https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-treatment-multiple-sclerosis_en-0.pdf
- European Medicines Agency. Qualification opinion of clinically interpretable treatment effect measures based on recurrent event endpoints that allow for efficient statistical analyses, https://www.ema.europa.eu/en/documents/other/qualification-opinion-clinically-interpretable-treatment-effect-measures-based-recurrent-event_en.pdf
- Naegelin Y, Naegelin P, von Felten S, et al. Association of rituximab treatment with disability progression among patients with secondary progressive multiple sclerosis. *JAMA Neurol* 2019; 76: 274–281.
- Lizak N, Malpas CB, Sharmin S, et al.; MSBase Study Group. Association of sustained immunotherapy with disability outcomes in patients with active secondary progressive multiple sclerosis. *JAMA Neurol* 2020; 77: 1398–1407.
- Lorscheider J, Jokubaitis VG, Spelman T, et al.; MSBase Study Group. Anti-inflammatory disease-modifying treatment and short-term disability progression in SPMS. *Neurology* 2017; 89: 1050–1059.
- Lorscheider J, Kuhle J, Izquierdo G, et al.; MSBase Study Group. Anti-inflammatory disease-modifying treatment and disability progression in primary progressive multiple sclerosis: A cohort study. *Eur J Neurol* 2019; 26: 363–370.
- Hauser SL, Bar-Or A, Comi G, et al. Ocrelizumab versus interferon beta-1a in relapsing multiple sclerosis. *N Engl J Med* 2017; 376: 221–234.
- Montalban X, Hauser SL, Kappos L, et al. Ocrelizumab versus placebo in primary progressive multiple sclerosis. *N Engl J Med* 2017; 376: 209–220.
- Lin D, Wei L, Yang I, et al. Semiparametric regression for the mean and rate functions of recurrent events. *J R Stat Soc Ser B Stat Methodol* 2000; 62: 711–730.
- Cook RJ and Lawless JF. *The statistical analysis of recurrent events*. New York: Springer Science & Business Media, 2007, pp. 128–133.
- Bühler A. *Comparison of time-to-first-event and recurrent event methods in multiple sclerosis trials*. Master's Thesis, Ulm University, Ulm, 2019, <http://arxiv.org/abs/2111.01937>
- Wang YC, Meyerson L, Tang YQ, et al. Statistical methods for the analysis of relapse data in MS clinical trials. *J Neurol Sci* 2009; 285: 206–211.
- Mütze T, Glimm E, Schmidli H, et al. Group sequential designs for negative binomial outcomes. *Stat Methods Med Res* 2019; 28(8): 2326–2347.
- Kappos L, Wolinsky JS, Giovannoni G, et al. Contribution of relapse-independent progression vs relapse-associated worsening to overall confirmed disability accumulation in typical relapsing multiple sclerosis in a pooled analysis of 2 randomized clinical trials. *JAMA Neurol* 2020; 77: 1132–1140.
- Schmidli H, Roger JH and Akacha M. Estimands for recurrent event endpoints in the presence of a terminal event. *Stat in Biopharm Rese*. Epub ahead of print 12 October 2021. DOI: 10.1080/19466315.2021.1895883.
- Furberg JK, Rasmussen S, Andersen PK, et al. Methodological challenges in the analysis of recurrent events for randomised controlled trials with application to cardiovascular events in LEADER. *Pharm Stat* 2022; 21(1): 241–267.
- Koch MW, Mostert J, Repovic P, et al. Reliability of outcome measures in clinical trials in secondary progressive multiple sclerosis. *Neurology* 2021; 96(1): e111–e120.
- Fox EJ, Markowitz C, Applebee A, et al. Ocrelizumab reduces progression of upper extremity impairment in patients with primary progressive multiple sclerosis: Findings from the phase III randomized ORATORIO trial. *Mult Scler* 2018; 24(14): 1862–1870.

Visit SAGE journals online
[journals.sagepub.com/
home/msj](https://journals.sagepub.com/home/msj)

 SAGE journals