abbvie

Composite endpoints to measure disability progression or improvement in patients with multiple sclerosis

- Jun Zhao, Weining Robieson, Adam Ziemann, and George Haig

Regulatory-Industry Statistics Workshop

Washington DC, September 2018

Disclosure

This presentation was sponsored by AbbVie. AbbVie contributed to writing, reviewing, and approving the publication.

Jun Zhao, Weining Robieson, Adam Ziemann, and George Haig are employees of AbbVie, Inc.

Highlights

In this talk, we describe the composite clinical endpoints that measure the disability progression or disability improvement (focus on improvement) in subjects with multiple sclerosis (MS).

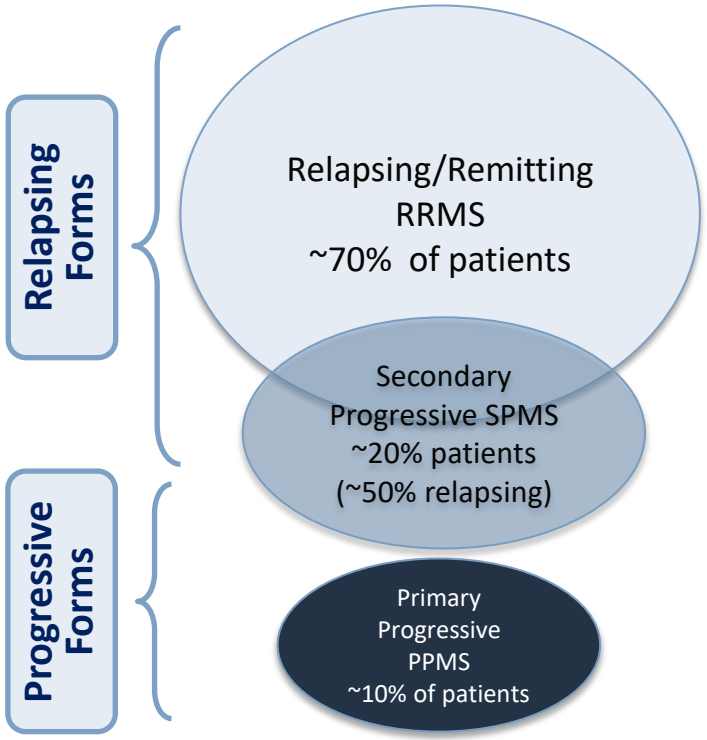
1) Multiple instruments can measure disability in functional improvement or functional worsening, including Timed 25-Foot Walk (T25FW), 9-Hole Peg Test (9HPT) and Expanded Disability Status Scale (EDSS).
2) The current definitions of disability in functional improving or worsening for each domain are accepted and used in the medical field.
3) The composite endpoint (EDSS+) may be more sensitive to detect disability improvement or disability progression than using any single instrument alone. Using a publicly available database, we explored the distribution and time course of the events in MS patient subtypes.
4) The shortfall of applying the composite endpoint EDSS+ is that the treatment duration might be long (e.g., 2-3 years) in order to collect enough events. We are considering an endpoint, Overall Response Score (ORS), that may be more sensitive to detect a treatment signal in a short treatment duration study (e.g., ½-1 year). Using the publicly available database, we explored the correlation between these endpoints.
5) There are some statistical challenges to use these endpoints.

Outline

- Brief background of the disease
- Construct composite endpoints based on individual domain scores
- The relationship between the endpoints and relative strengths
- Statistical challenges
- Remarks

Brief background

- Multiple sclerosis (MS) is a chronic autoimmune and neurodegenerative disorder of the central nervous system (CNS) that is characterized by inflammation, demyelination, axonal transection, and neuronal loss.



Source: Decision Resources Patient Base (2015) – excludes Clinically Isolated Syndrome (CIS)

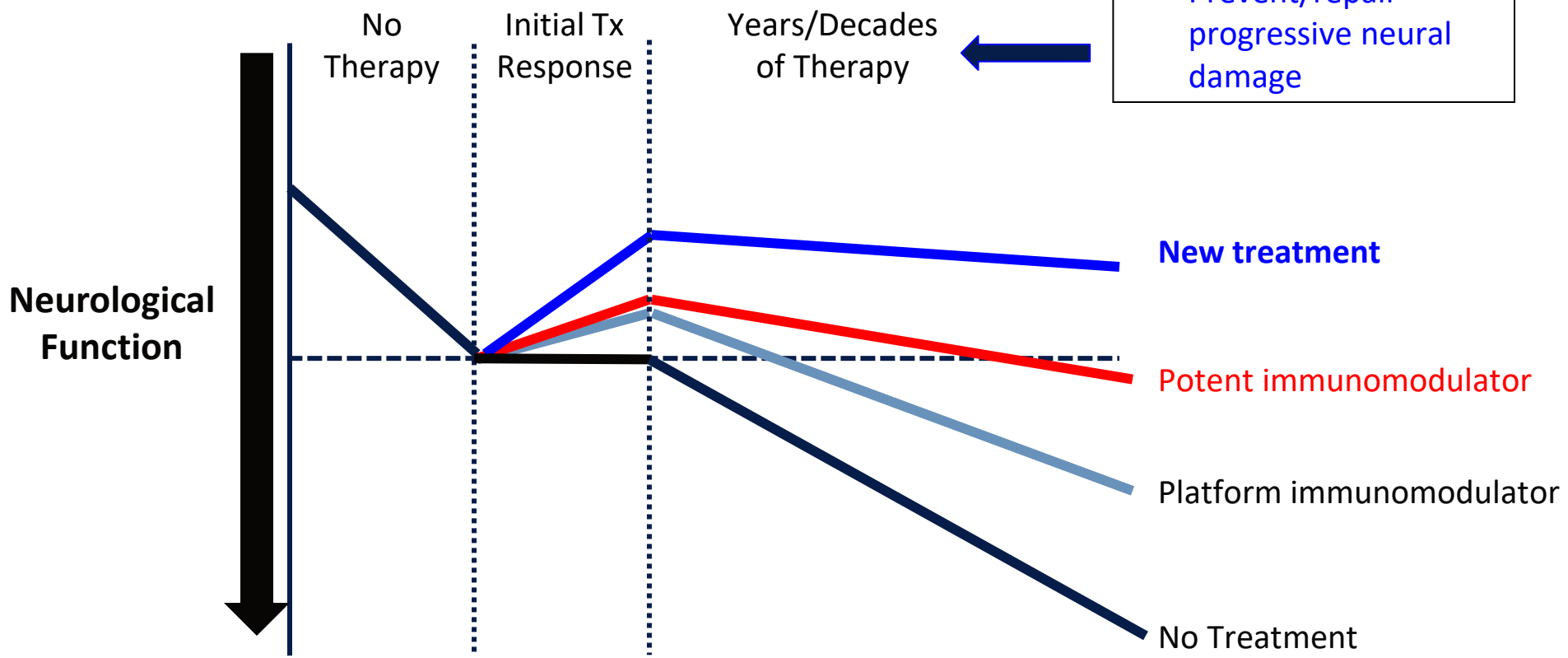
- In spite of therapeutic advances, many patients with MS continue to experience disability progression, with or independent of relapses, with a concordant loss of physical and cognitive function.

MS treatment goals on neurological function

Initial MS Tx Goals: Resolve / Prevent / Repair Prior Damage

Long-Term MS Tx Goals:

- Prevent/repair new relapses
- Prevent/repair progressive neural damage



Measuring MS disability – EDSS, MSFC (T25FW, 9HPT)

- MS disability is difficult to measure. Affected domains include motor, sensory, coordination, visual, cognitive, and psychological.

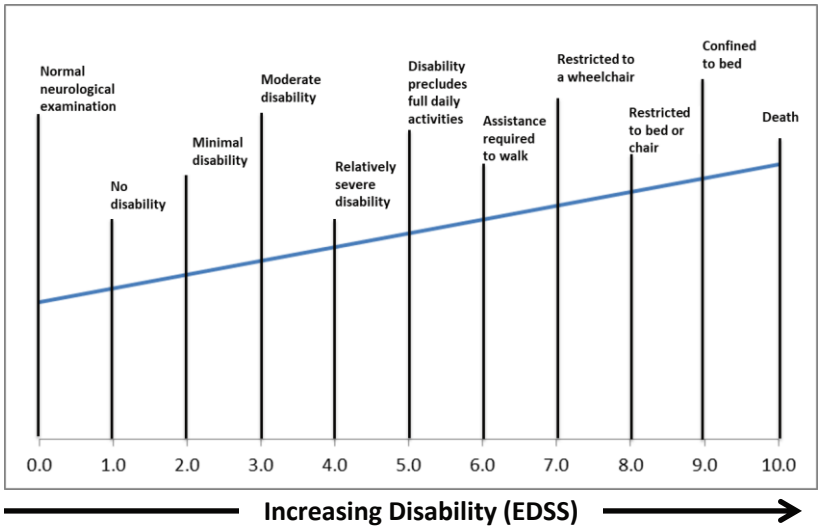
- EDSS (The Expanded Disability Status Scale) is a composite score designed to quantify disability based on scores of eight distinct functional systems.

  Limitations of the EDSS in assessing disability in MS are well known:

  Scoring subjectivity, resulting in poor reliability within and between raters;

  Insensitive to detecting changes in many aspects of function;

  It is a non-linear ordinal scale: Populations show a bimodal distribution of EDSS categories, and the rate of progression through the scale varies by baseline score



[*Reference*] Kurtzke JF.Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS).Neurology 1983;33(11):1444-1452.

- MSFC (The Multiple Sclerosis Functional *Composite*)

  Ambulation: timed 25-foot walk (**T25FW)**, time to completion (seconds) of walking over a distance of 25 ft

  Upper-extremity: 9-hole peg test **(9HPT)**, time to completion (seconds)

  Cognition: paced auditory serial addition test (**PASAT),** number of correct response (of total of 60)

[*Reference*] Cutter GR, Baier ML, Rudick RA, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain 1999; 122: 871–82.

Measuring functional disability

- Four key functional domains Should be considered:
  - Time to ambulation
  - Manual dexterity
  - Visual acuity
  - Cognition
- Need to have validated assessment tools and clinically meaningful change thresholds for each domain.
- Open database: Placebo arms from clinical trial datasets, which were contributed by industry members of MSOAC*, are aggregated in the MSOAC Placebo Database. The MSOAC Placebo Database presently includes 2465 individual patient records from 9 clinical trials. The current version 1.0 includes records from relapsing-remitting, secondary progressive, and primary progressive forms of MS.

Assessment	Domain	Progression Threshold	Improvement Threshold
EDSS	Multiple	1 point (severity) ⬆	1 point (severity) ⬇
T25FW	Walking Speed	20% more time [1,2] ⬆	20% less time [1] ⬇
9HPT	Manual Dexterity	20% more time [2] ⬆	20% less time [2] ⬇

[*Reference*] [1]Hobart J, Blight AR, Goodman A, et al. Timed 25-foot walk: direct evidence that improving 20% or greater is clinically meaningful in MS. Neurology. 2013;80(16):1509-17.
[*Reference*] [2]Kragt J, Van der Linden E, Nielsen J, Ultdehaag B, Polman CH. Clinical impact of 20% worsening on Timed 25-foot Walk and 9-hole Peg Test in multiple sclerosis. Mult Scler 2006; 12:594-598.
* The Multiple Sclerosis Outcome Assessments Consortium (MSOAC), a coalition of the National MS Society, industry, academia, patient representatives, FDA, EMA, and the Critical Path Institute.

Use a composite endpoint **(EDSS+)** that incorporates: T25FW, 9HPT, & EDSS

EDSS Progression (or Improvement)
T25FW Progression (or Improvement)
9HPT Progression (or Improvement)

A patient that has sustained progression (or improvement) in ***any*** of these domains is considered a "responder"

The observed disability progression (or improvement) need to be confirmed 12 weeks or 24 weeks later

Improvement

- ≥ 20% decrease from baseline for 9HPT (Upper extremity dexterity), or
- ≥ 20% decrease from baseline for 9HPT T25FW (Lower extremity ambulation) , or
- ≥ 1.0 point decrease from baseline EDSS score of 2.0 or greater

Confirmation

Confirmed improvement is measured as either a 12-week or as a 24-week sustained change from baseline

Patient Inclusion

Recruit patients with some minimum level of definite functional disability: EDSS ≥2.0

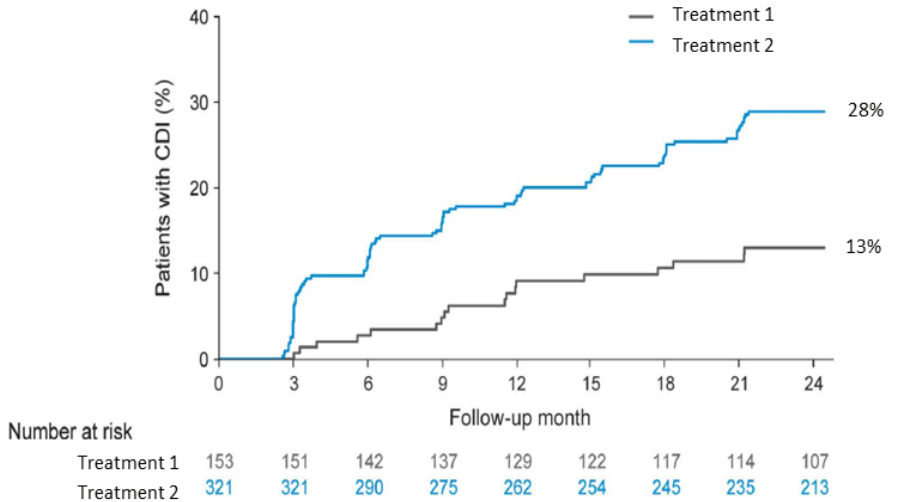


Figure is for Illustration purpose

EDSS+ more sensitive than each individual domain

- Recent research [Reference] has suggested that a combined disability measure that incorporates scoring from multiple instruments, for example, using T25FW, 9HPT, and EDSS, may be more sensitive to detect associated disability progression or disability improvement events than with any single instrument alone.

Advantages

- Behavior of individual and composite endpoints well understood for assessing disease *progression*; utility for assessing *improvement* less clear
- Increasingly used in Relapsing MS and Progressive MS trials
- Attractive to HCPs and payers

Disadvantages

- Dichotomous endpoint; full magnitude of progression or improvement not captured
- Likely requires long treatment (e.g., 24 months + 6 months confirmation) to detect group differences
- For improvement: does not account for improvement across multiple domains, or improvement in some with worsening in others

[*Reference*] Diego Cadavid, Jeffrey A Cohen, et al. EDSS-Plus, an improved endpoint for disability progression in secondary progressive multiple sclerosis. Multiple Sclerosis Journal, 2017, Vol 23(1) 94-105.

Statistical considerations of using EDSS+

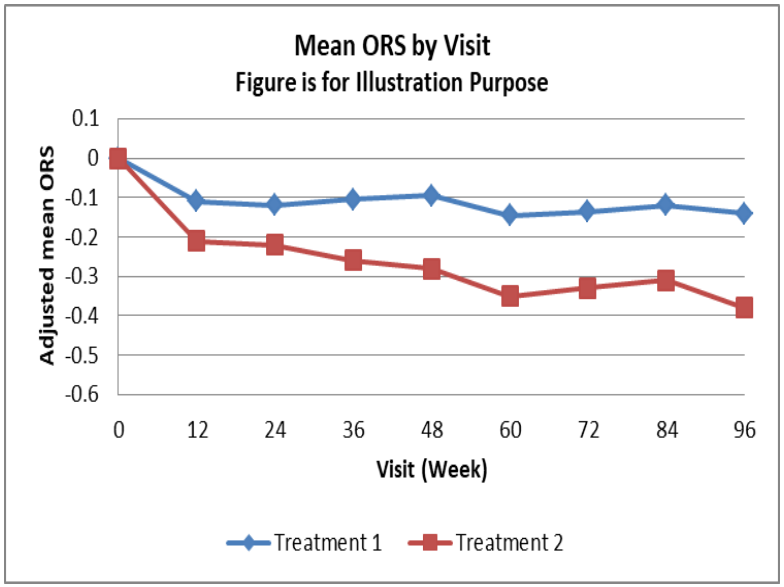
- Need a large and long treatment duration study to detect treatment difference. Both the accumulation of events and the confirmation from a later visit contribute to the lengthy duration
- Challenge in population selection. E.g., for detecting disability improvement, patients should have at least one domain deficient
- Longitudinal collected observations that may be spaced by at least 12-weeks for all three domains (for time to event data)
- Confounding factors may impact the detection of disability progression /improvement, e.g., relapse
- Handling of contradiction among domains (one domain improving, another worsening)
- Relatively high drop-out rate was observed in historical data, e.g., 10% per year. In many cases, when an event was onset, it may not be confirmed due to missing data. One research explored multiple analysis/imputation methods when informative missing is expected. [Reference]

[Reference] Zhao, J, Tang, Q, Fu, B, Pan, Q, and Tsao, C. 2015. Methods of Assessing Treatment Failure or Response with Informative Censoring. In *JSM Proceedings*, Biopharmaceutical Section. Alexandria, VA: American Statistical Association. 3846 - 3853.

New composite endpoint: construct improvement or worsening in *each* domain into a single non-dichotomous score: the **ORS** (overall response score)

ORS algorithm:

- Clinically significant improvement from baseline generates +1 point for *each* domain (EDSS, T25FW, 9HPT-Right hand, 9HPT-Left habnd: same thresholds as EDSS+ endpoint)
- Clinically significant worsening from baseline generates -1 point for each endpoint (same thresholds as improvement in opposite direction)
- Summed scores are analyzed in a continuous manner (possible range of -4 to +4 for each visit)



Statistical considerations of using ORS

Pros/Cons

- The ORS may be a more sensitive endpoint to enable signal detection with shorter treatment duration
- Take both progression in one domain and improvement in another domain into account
- Publication of these data and validation against anchored functional scales are limited
- The equal weight of the 4 domains in the composite score may be questionable
- The ORS doesn't need confirmation, which may increase variability for the patient population

Statistical Challenges

- An MMRM with covariance adjustment may be used to estimate different contrasts of treatment and visit
- The magnitude of the improvement/worsening is not used, and the ceiling effect of the score may impact detecting the treatment difference
- Confounding factors may impact the detection of disability improvement/progression , e.g., relapse during the study
- Challenge on how to handle one/two domain missing (partial data missing)
- ORS may be suitable for a Phase 2 proof-of-concept (POC) study. For confirmative studies, primary endpoint may need to be EDSS+ (need to have regulatory agreement).

Correlation between the two endpoints, and mapping the required response on EDSS+ to ORS

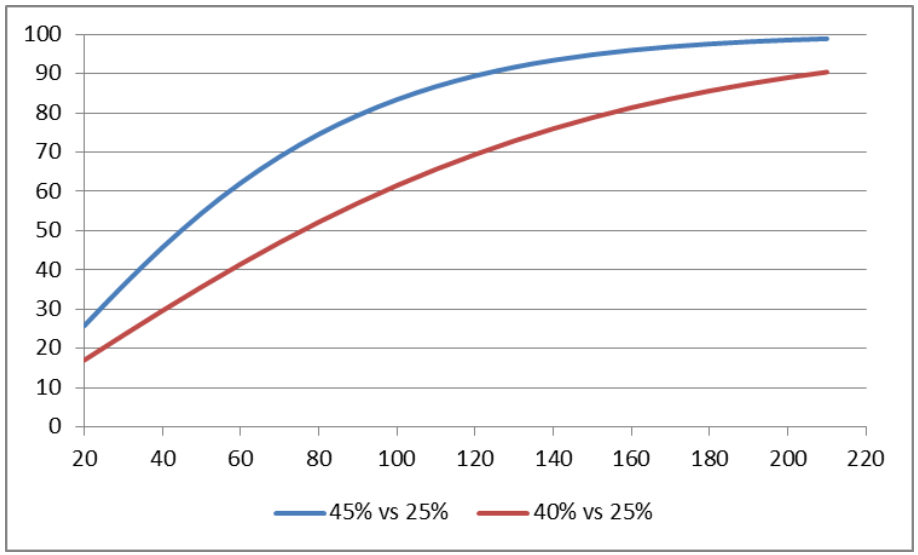
- Algorithms to find the correlation between the EDSS+ and ORS endpoints
  1) Use an open database , e.g., MSOAC (placebo patients)
  2) Select a population (the following are for illustration purpose)
     e.g., subgroup of patients with at least one domain deficiency
     e.g., subgroup with baseline EDSS>=2.0 (no imputation)
  3) Model fitting using the open database:
     - ❑ Logistic regression on event: event ~ ORS at Week 48 + covariates
       e.g., logit(event) = -2.1 + 1.1 x, OR =3.0, Concordance = 73.1
     - ❑ Time to event ~ ORS at Week 24 + covariates

- Estimate placebo effect
  - EDSS+: e.g., 12% - 20% overall event rate (no imputation)
  - ORS: mean and variance: e.g., -0.39 (1.00)

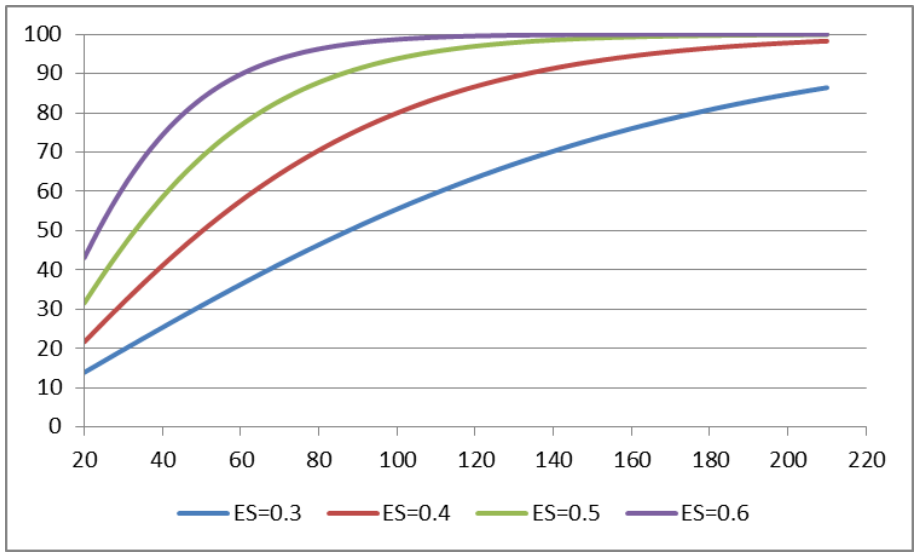
- Simulate corresponding effect size between groups
  - EDSS+: e.g., 25% vs 45% event rate
  - ORS: effect size

Example of power vs sample size by using EDSS+ and ORS

Power vs Sample Size by using EDSS+



Power vs Sample Size by using ORS



nQuery (Log-rank Test of survival in two groups followed for fixed time, constant hazard ratio) is used for time to event data.
Two-sample t-test with equal variance is used for continuous data.
Alpha=0.05 two sided, no missing data adjusted for sample size calculation.

Application: simulating adaptive clinical studies with a surrogate endpoint for interim decision-making

- When it takes a long time to observe the primary endpoint, e.g., time-to-event endpoint with long follow-up, utilizing information on an early surrogate endpoint at an interim analysis (IA) is more efficient for selecting which dose to continue.
- The correlation between the primary endpoint and the surrogate endpoint can be incorporated properly to evaluate dose groups at IA.
- Incorporation of correlated surrogate (e.g., biomarkers) endpoints is supported by some statistical software, e.g. ADDPlan, but may require assumptions of concordance between the primary endpoint and the surrogate endpoints. Authors found out that specifying a concordance between the two endpoints, i.e., how often they agree in terms of making the target delta, dominates over the specification of the correlation between the two endpoints. [Reference 1]
- The authors proposed a Bayesian model-based approach for simulating clinical trials using a surrogate endpoint for treatment selection at interim decision-making, and compared with standard parallel designs. Sample size savings depend on the enrollment rates and the timing of the interim analysis. [Reference 2]

[Reference 1] Ziqian Geng, Bo Fu, Alan Hartford, and Jun Zhao. Risk analysis on using surrogate endpoint at interim analysis. Presentation in the JSM, 2016.
[Reference 2] Xiaotian Chen, Alan Hartford, Mei Li, Jun Zhao. A Model-Based Approach for Simulating Adaptive Clinical Studies with Surrogate Endpoints Used for Interim Decision-Making. Presentation in the JSM, 2018

Concluding Remarks

- Composite endpoints could be binary, time to event (e.g. EDSS+) or continuous (categorical that could be analyzed as continuous, e.g. ORS)
- ORS overcomes some challenges that the 12-week or 24-week confirmed disability progression / improvement (EDSS+), and may be a suitable endpoint for POC studies
- Mapping between binary and continuous composite endpoints could be explored using existing database to inform late phase study design.