The importance of informative visit patterns

Thomas Debray

@TPA_Debray

Acknowledgements

Massimiliano Copetti[1], Robert W. Platt[2,3], Gabrielle Simoneau[2,4,*], Fabio Pellegrini[5], Carl de Moor [6]

This work was supported by Biogen (Cambridge, MA, USA).

[*] Several slides are based on materials kindly provided by Gabrielle Simoneau

[1] Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, FG, Italy; [2] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada; [3] Centre for Clinical Epidemiology, Jewish General Hospital, Montreal, QC, Canada; [4] Biogen Canada, Mississauga, ON, Canada; [5] Biogen International GmbH, Zug, Baar, Switzerland; [6] Biogen, Cambridge, MA, USA

Background

Randomized trials are commonly used to assess relative treatment effects

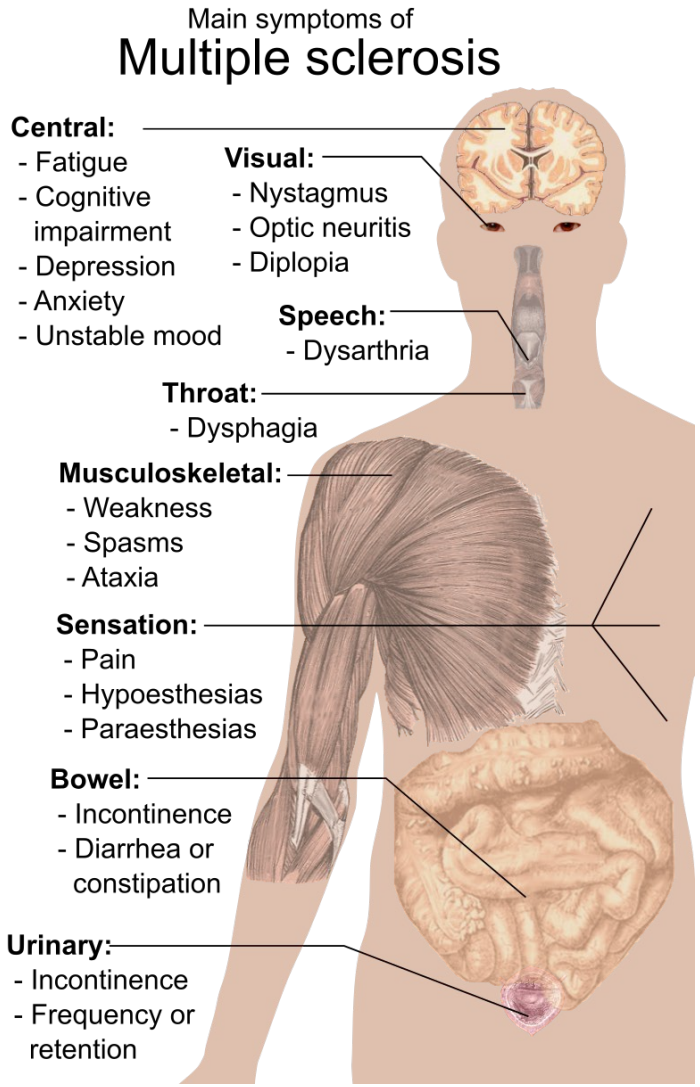
Non-randomized data sources may help to

- study effectiveness of therapeutic interventions in less controlled environments
- evaluate prognosis of individual patients encountered in routine care
- understand variations in treatment and outcomes
- examine factors that influence prognosis and quality of life
- describe care patterns
- monitor safety and harm

Multiple Sclerosis

Multiple sclerosis (MS) is a chronic progressive disorder that affects approximately 2.3 million people worldwide

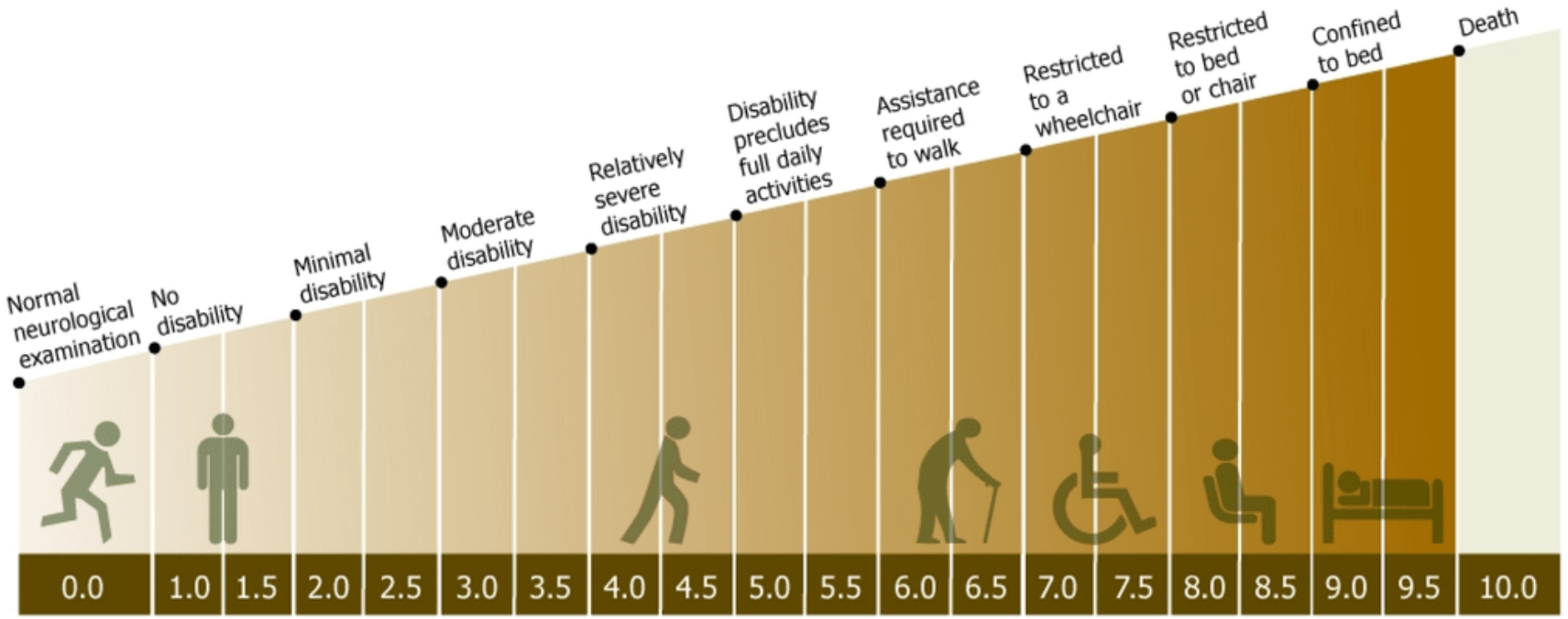
- Most patients diagnosed with MS have a relapsing-remitting form of the disease
- Relapse-remitting MS is characterized by episodes of disease activity (relapses) when symptoms get worse
- No known cure
- 16 disease modifying therapies (DMTs) available



Quantifying disease severity

Expanded Disability Status Scale (EDSS)

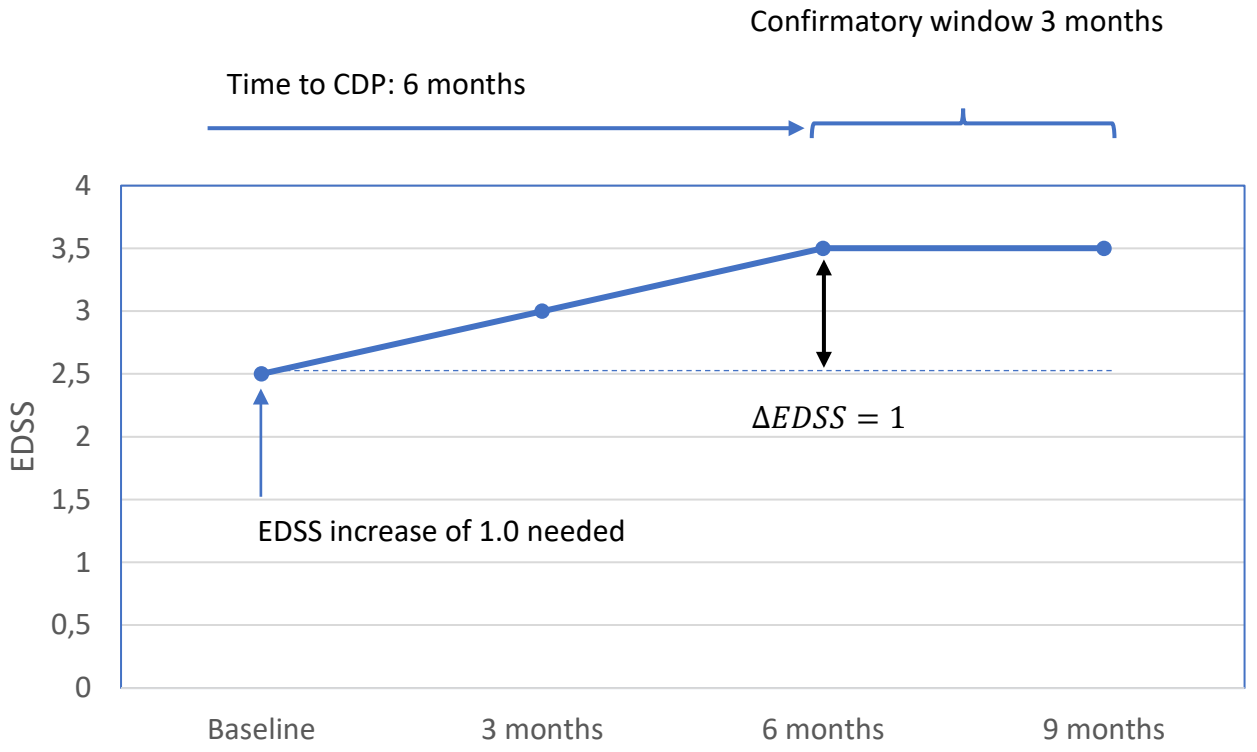
- Semi-continuous scale
- Ranges from 0 (normal function) to 10 (death) by increments of 0.5 points.



Source image: https://mamametms.nl/edss/

Time to Confirmed Disease Progression

- Confirmed disability progression (CDP) is commonly used as one of the efficacy endpoints in randomized controlled trials (RCTs).
- Calculation of CDP requires standardized follow-up visits with measurement of EDSS (e.g. assessment every 3 months)
- Time from baseline to an EDSS increase of:
  - ≥ 1.5 points if baseline EDSS = 0;
  - ≥ 1.0 point if baseline EDSS < 6.0;
  - ≥ 0.5 point if baseline EDSS ≥ 6.0;
  - Increase must be confirmed 3 months later



MS registries

- Record information about the health status of a patient when they visit their doctor – no planned visit schedule
- May be hospital-based, country-based or multi-national

Source images: https://www.rigshospitalet.dk/, https://www.msbase.org/, http://www.narcrms.org/, http://www.ofsep.org/, https://www.va.gov/ms/

Challenges in non-randomized data sources

Table 1 | Bias domains included in ROBINS-I

Domain	Explanation
Pre-intervention	**Risk of bias assessment is mainly distinct from assessments of randomised trials**
Bias due to confounding	Baseline confounding occurs when one or more prognostic variables (factors that predict the outcome of interest) also predicts the intervention received at baseline ROBINS-I can also address time-varying confounding, which occurs when individuals switch between the interventions being compared and when post-baseline prognostic factors affect the intervention received after baseline
Bias in selection of participants into the study	When exclusion of some eligible participants, or the initial follow-up time of some participants, or some outcome events is related to both intervention and outcome, there will be an association between interventions and outcome even if the effects of the interventions are identical This form of selection bias is distinct from confounding—A specific example is bias due to the inclusion of prevalent users, rather than new users, of an intervention
At intervention	**Risk of bias assessment is mainly distinct from assessments of randomised trials**
Bias in classification of interventions	Bias introduced by either differential or non-differential misclassification of intervention status Non-differential misclassification is unrelated to the outcome and will usually bias the estimated effect of intervention towards the null Differential misclassification occurs when misclassification of intervention status is related to the outcome or the risk of the outcome, and is likely to lead to bias
Post-intervention	**Risk of bias assessment has substantial overlap with assessments of randomised trials**
Bias due to deviations from intended interventions	Bias that arises when there are systematic differences between experimental intervention and comparator groups in the care provided, which represent a deviation from the intended intervention(s) Assessment of bias in this domain will depend on the type of effect of interest (either the effect of assignment to intervention or the effect of starting and adhering to intervention).
Bias due to missing data	Bias that arises when later follow-up is missing for individuals initially included and followed (such as differential loss to follow-up that is affected by prognostic factors); bias due to exclusion of individuals with missing information about intervention status or other variables such as confounders
Bias in measurement of outcomes	Bias introduced by either differential or non-differential errors in measurement of outcome data. Such bias can arise when outcome assessors are aware of intervention status, if different methods are used to assess outcomes in different intervention groups, or if measurement errors are related to intervention status or effects
Bias in selection of the reported result	Selective reporting of results in a way that depends on the findings and prevents the estimate from being included in a meta-analysis (or other synthesis)

Challenges in non-randomized data sources

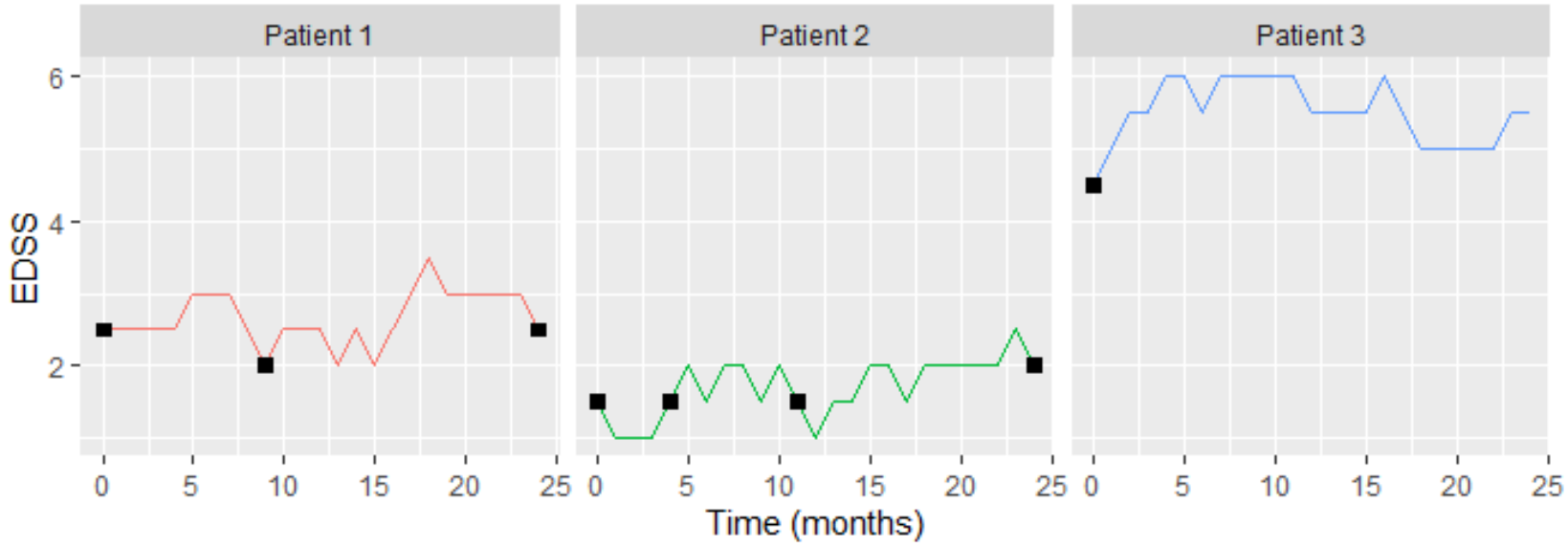
Table 1 | Bias domains included in ROBINS-I

Domain	Explanation
Pre-intervention	**Risk of bias assessment is mainly distinct from assessments of randomised trials**
Bias due to confounding	Baseline confounding occurs when one or more prognostic variables (factors that predict the outcome of interest) also predicts the intervention received at baseline ROBINS-I can also address time-varying confounding, which occurs when individuals switch between the interventions being compared and when post-baseline prognostic factors affect the intervention received after baseline
Bias in selection of participants into the study	When exclusion of some eligible participants, or the initial follow-up time of some participants, or some outcome events is related to both intervention and outcome, there will be an association between interventions and outcome even if the effects of the interventions are identical This form of selection bias is distinct from confounding—A specific example is bias due to the inclusion of prevalent users, rather than new users, of an intervention
At intervention	**Risk of bias assessment is mainly distinct from assessments of randomised trials**
Bias in classification of interventions	Bias introduced by either differential or non-differential misclassification of intervention status Non-differential misclassification is unrelated to the outcome and will usually bias the estimated effect of intervention towards the null Differential misclassification occurs when misclassification of intervention status is related to the outcome or the risk of the outcome, and is likely to lead to bias
Post-intervention	**Risk of bias assessment has substantial overlap with assessments of randomised trials**
Bias due to deviations from intended interventions	Bias that arises when there are systematic differences between experimental intervention and comparator groups in the care provided, which represent a deviation from the intended intervention(s) Assessment of bias in this domain will depend on the type of effect of interest (either the effect of assignment to intervention or the effect of starting and adhering to intervention).
Bias due to missing data	Bias that arises when later follow-up is missing for individuals initially included and followed (such as differential loss to follow-up that is affected by prognostic factors); bias due to exclusion of individuals with missing information about intervention status or other variables such as confounders
Bias in measurement of outcomes	Bias introduced by either differential or non-differential errors in measurement of outcome data. Such bias can arise when outcome assessors are aware of intervention status, if different methods are used to assess outcomes in different intervention groups, or if measurement errors are related to intervention status or effects
Bias in selection of the reported result	Selective reporting of results in a way that depends on the findings and prevents the estimate from being included in a meta-analysis (or other synthesis)

Challenges in non-randomized data sources

Patient visits do not occur according to a predefined schedule (e.g. every 3 months) but are dictated by patient characteristics or treatment choices

-> Observed sequences of disease scores may not adequately reflect disease progression



Impact of visit frequency in MS

Patient visits are dictated by patient characteristics or treatment choices:

- Changes in disease severity (e.g. relapses) may affect visit frequency
- Different hospitals/clinics/physicians could encourage different visit schedules
- Different DMTs may require different monitoring schedules

Informative visit patterns occur when reasons associated with visit frequency are also predictors of the outcome.

	Alemtuzumab (n=156)	Interferon beta (n=282)
Women	110 (71%)	209 (74%)
Age, years	33 (8)	33 (9)
Disease duration, years	3·1 (1·9 to 6)	2·8 (1·3 to 6·5)
Relapses in 12 months before baseline	2 (1·3)	2·1 (1)
Disability, EDSS step	3·0 (2·0 to 4·0)	3·0 (2·0 to 4·0)
Difference between baseline date and the date of baseline EDSS measurement	0 (–38 to 13)	–15 (–51 to 0)
Intervisit interval, months	9 (6 to 12)	3 (1 to 5)
Previous therapies	0 (0 to 1)	0 (0 to 1)
Most active previous therapy		
Interferon beta or glatiramer acetate	45 (29%)	87 (31%)
Teriflunomide	0	0
Dimethyl fumarate	0	0
Fingolimod	0	0
Natalizumab	2 (1%)	2 (1%)
Mitoxantrone	2 (1%)	6 (2%)
Other	0	0
No previous treatment	102 (65%)	183 (65%)
Length of pairwise-censored follow-up on study therapy, years	2·1 (1·0 to 3·9)	2·1 (1·0 to 3·9)

Table 1 in Kalincik T, et al. (2017). Treatment effectiveness of alemtuzumab compared with natalizumab, fingolimod, and interferon beta in relapsing-remitting multiple sclerosis: a cohort study. *Lancet Neurology*.

Impact of visit frequency in MS

Calculation of CDP is highly problematic

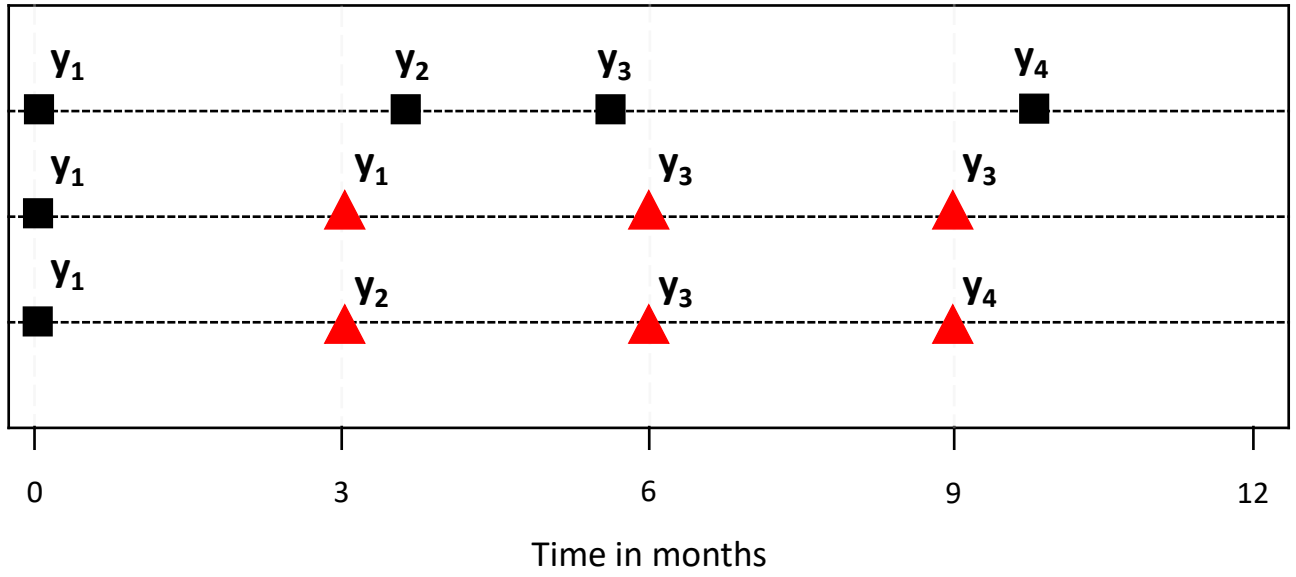
- Irregular visit patterns may lead to outcome assessment bias due to informative missingness of relevant patient outcomes.
- Irregular visit patterns may lead to participant selection bias due to exclusion of participants with no follow-up visits

Existing mitigation strategies for unequal follow-up visit patterns

Naïve methods

General idea: map the irregular visit pattern to an equally spaced visit schedule appropriate for applying the definition of time to CDP

- Last observation carried forward (LOCF): replaces the missing response by the patient's most recent observation.
- Rounding: replaces the missing patient's response by their nearest observation, past or future.



Pitfalls of LOCF

LOCF is only unbiased when

- The missing data are MCAR; and
- The data used as the basis for the LOCF imputation has exactly the same distribution as the unknown missing data.

Bias due to LOCF is often assumed to lead to an underestimation of treatment effects.

However, LOCF analyses can introduce a **positive or negative bias**

Ref: Lachin JM. Fallacies of last observation carried forward analyses. Clin Trials. 2016;13(2).

Pitfalls of rounding

Patient ID	Treatment	Months	Rounded months	EDSS	Progression
1	A	0	0	1	0
1	A	4.4	3	2	1
1	A	7	6	2	.
1	A	10	9	3	.
2	B	0	0	1	0
2	B	2	3	1	0
2	B	4.6	6	2	1
2	B	9	9	2.5	.

Reconstructed time to CDP: 3 months

Reconstructed time to CDP: 6 months

- Rounding artificially created a 3-month lag between time to CDP of the two patients.
- Conclusion: treatment B slows time to CDP compared to treatment A when the progressions were recorded only 6 days apart (0.2 month).

Need for more advanced methods

Key issues

- LOCF and rounding intend to replace missing values by a *plausible* value (imputation)
- LOCF and rounding do not account for uncertainty in imputed values
- Non-random visit patterns will typically lead to bias if imputation does not adjust for the underlying missing data mechanism
- More accurate imputations do not necessarily lead to less bias in treatment effect estimates

"The goal of MI is not to estimate the missing values themselves but rather to produce unbiased and efficient estimates for the population parameters of interest, by essentially averaging over the (unknown) distribution of the missing data"

Ref: Perkins NJ, Cole SR, Harel O, Tchetgen Tchetgen EJ, Sun B, Mitchell EM, et al. Principled Approaches to Missing Data in Epidemiologic Studies. Am J Epidemiol. 2018 01;187(3):568–75.

Imputations from Generalized Linear Mixed Models

Proposed multi-level modeling

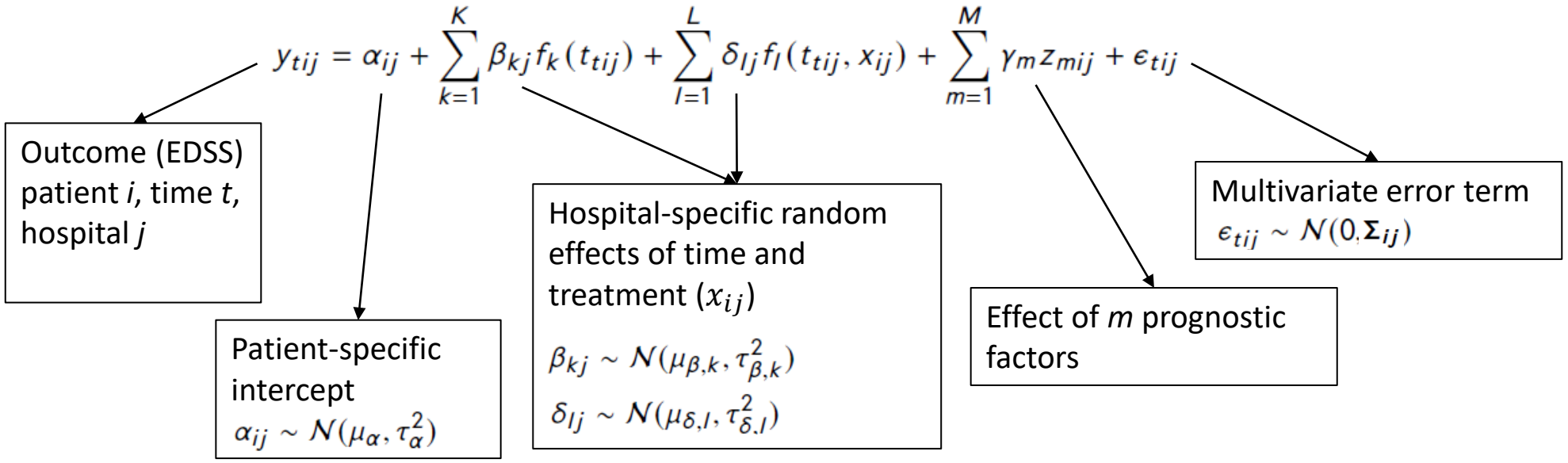
We propose to model the EDSS trajectories of individual patients with linear mixed models and to generate imputations from the fitted model

Key advantages:

- Do not require observations at fixed intervals, borrow strength across observations over time
- Account for the repeated measures within patient
- Can account for higher-level clustering (e.g. hospital, country)
- Possible to account for (time-varying) prognostic factors

Proposed multi-level modeling

Model the outcome over time with a linear mixed model with patient- and hospital-specific random effects.



Exponential spatial correlation structure (Σ_{ij}) to account for within-patient correlations.

Generating imputed values

Single imputation

The expected value for a missing EDSS score at time $t_{\varphi ij}$ is

$$\hat{y}_{\phi ij} = \hat{\alpha}_{ij} + \sum_{k=1}^{K} \hat{\beta}_{kj} f_k(t_{tij}) + \sum_{l=1}^{L} \hat{\delta}_{lj} f_l(t_{tij}, x_{ij}) + \sum_{m=1}^{M} \hat{\gamma}_m z_{mij} + \hat{e}_{\phi ij}$$

- In the absence of autocorrelation, the expected value of $y_{\varphi ij}$ is independent of the residual errors and we may set $\hat{e}_{\varphi ij} = 0$
- In the presence of autocorrelation, the observed residual errors $\boldsymbol{\varepsilon_{ij}} = \left(\varepsilon_{1ij}, \ldots, \varepsilon_{n_{ij}ij}\right)^{\boldsymbol{T}}$ may inform the magnitude of $\hat{e}_{\varphi ij}$. An improved prediction for $y_{\varphi ij}$ can thus be obtained by setting $\hat{e}_{\varphi ij} = E(e_{\varphi ij}|\boldsymbol{\varepsilon_{ij}})$
- The predictions $\hat{y}_{\varphi ij}$ are rounded to the nearest half-integer and truncated between 0 and 9.5.

Generating imputed values

Single imputation

Consider a patient with missing EDSS scores at times $s_{1ij}, \ldots, s_{m_{ij}ij}$ and observed EDSS scores at times $t_{1ij}, \ldots, t_{n_{ij}ij}$. The unobserved error terms can then be described by a multivariate normal distribution with conditional mean

$$\mathrm{E}\left(\begin{bmatrix} e_{s_{1ij}} \\ \vdots \\ e_{s_{m_{ij}ij}} \end{bmatrix} | \boldsymbol{\epsilon_{ij}}\right) = \boldsymbol{C_{ij}}\hat{\Sigma}_{ij}^{-1}\begin{bmatrix} \epsilon_{t_{1ij}} \\ \vdots \\ \epsilon_{t_{n_{ij}ij}} \end{bmatrix}$$

with

$$C_{ij} = \begin{bmatrix} \exp(-|s_{1ij} - t_{1ij}|/\hat{d})\hat{\sigma}^2 & \ldots & \exp(-|s_{1ij} - t_{n_{ij}ij}|/\hat{d})\hat{\sigma}^2 \\ \exp(-|s_{2ij} - t_{1ij}|/\hat{d})\hat{\sigma}^2 & \ldots & \exp(-|s_{2ij} - t_{n_{ij}ij}|/\hat{d})\hat{\sigma}^2 \\ \vdots & \ddots & \\ \exp(-|s_{m_{ij}ij} - t_{1ij}|/\hat{d})\hat{\sigma}^2 & \ldots & \exp(-|s_{m_{ij}ij} - t_{n_{ij}ij}|/\hat{d})\hat{\sigma}^2 \end{bmatrix}$$

However, imputing missing $y_{\varphi ij}$ by their expected value $\hat{y}_{\varphi ij}$ is problematic for statistical inference

Generating imputed values

Multiple imputation

We propose to randomly sample the residual terms from a conditional multivariate normal distribution:

$$\begin{pmatrix} \hat{e}_{s_{1ij}} \\ \vdots \\ \hat{e}_{s_{m_{ij}ij}} \end{pmatrix} \sim \text{MVN}\left(\text{E}\left(\begin{bmatrix} e_{s_{1ij}} \\ \vdots \\ e_{s_{m_{ij}ij}} \end{bmatrix} | \epsilon_{ij} \right), \text{Cov}\left(\begin{bmatrix} e_{s_{1ij}} \\ \vdots \\ e_{s_{m_{ij}ij}} \end{bmatrix} | \epsilon_{ij} \right) \right)$$

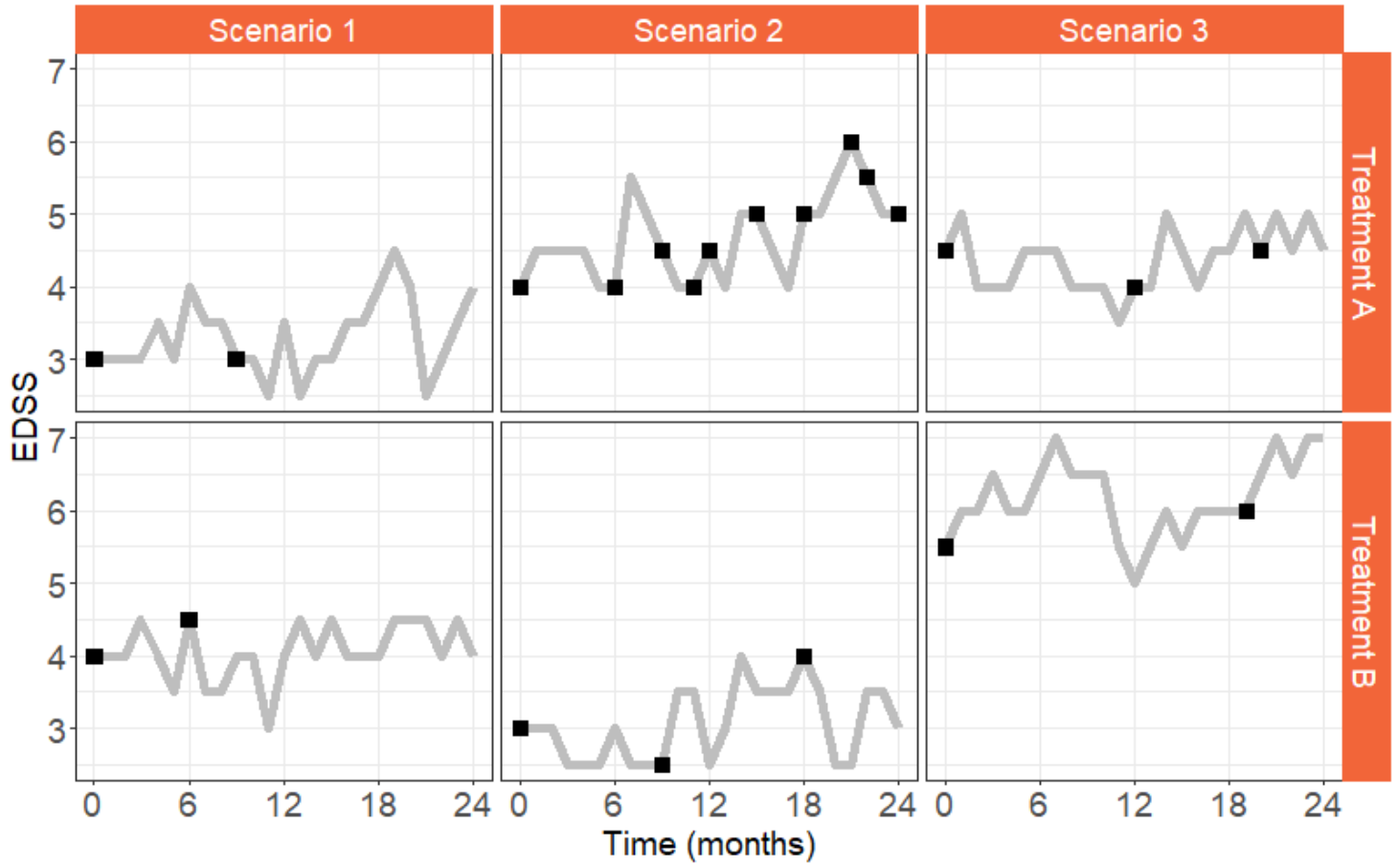
Ideally, additional noise is added to imputations to reflect uncertainty in estimated model parameters.

Simulation Study

Data Generation Mechanism

- 500 patients per hospital, 20 hospitals
- Treatment allocation as a function of age
- Visits are generated every month for a total follow-up of 24 months
- EDSS generated as an underlying continuous process from a linear mixed model:
  - Model parameter values based on published observational MS studies
  - Patient- and hospital-specific random intercepts
  - Age as a prognostic factor
  - Treatment contrast introduces a moderate treatment effect that accumulates over time, favoring treatment B.
  - AR1 correlation structure between monthly EDSS scores (rho = 0.8)
- Irregular visit patterns are introduced by deleting visits informatively
- 100 simulations per scenario

Data Generation Mechanism



	Scenario
1	Probability of observing a visit decreases from 36% at month 1 to 6% at month 24 (~ 4 visits / 24 months)
2	**Treatment A**: probability of observing a visit at 6, 12, 18 and 24 months is 85%, 3% otherwise (~ 5 visits/24 months) **Treatment B**: probability of observing a visit at 9 and 18 months is 67%, 3% otherwise (~ 3 visits/24 months)
3	Probability of observing a visit varies as a function of age, treatment and unobserved EDSS score (~ 4 visits/24 months)

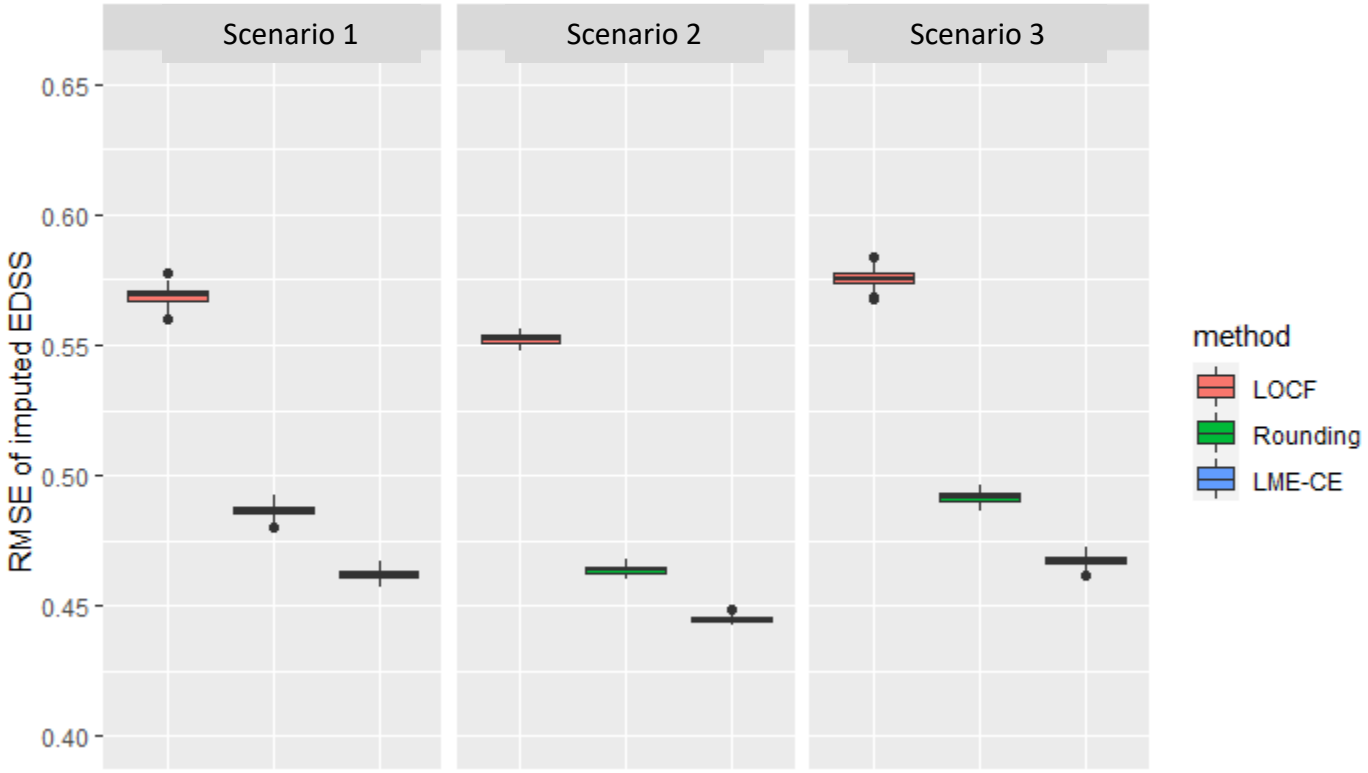
Imputation strategies

1. Last observation carried forward
2. Rounding
3. Proposed linear mixed model with 20 multiple imputations

The effect of treatment on time to CDP (with confirmation window of 3 months) is estimated with a Cox regression stratified by center and adjusted for age and baseline EDSS score.

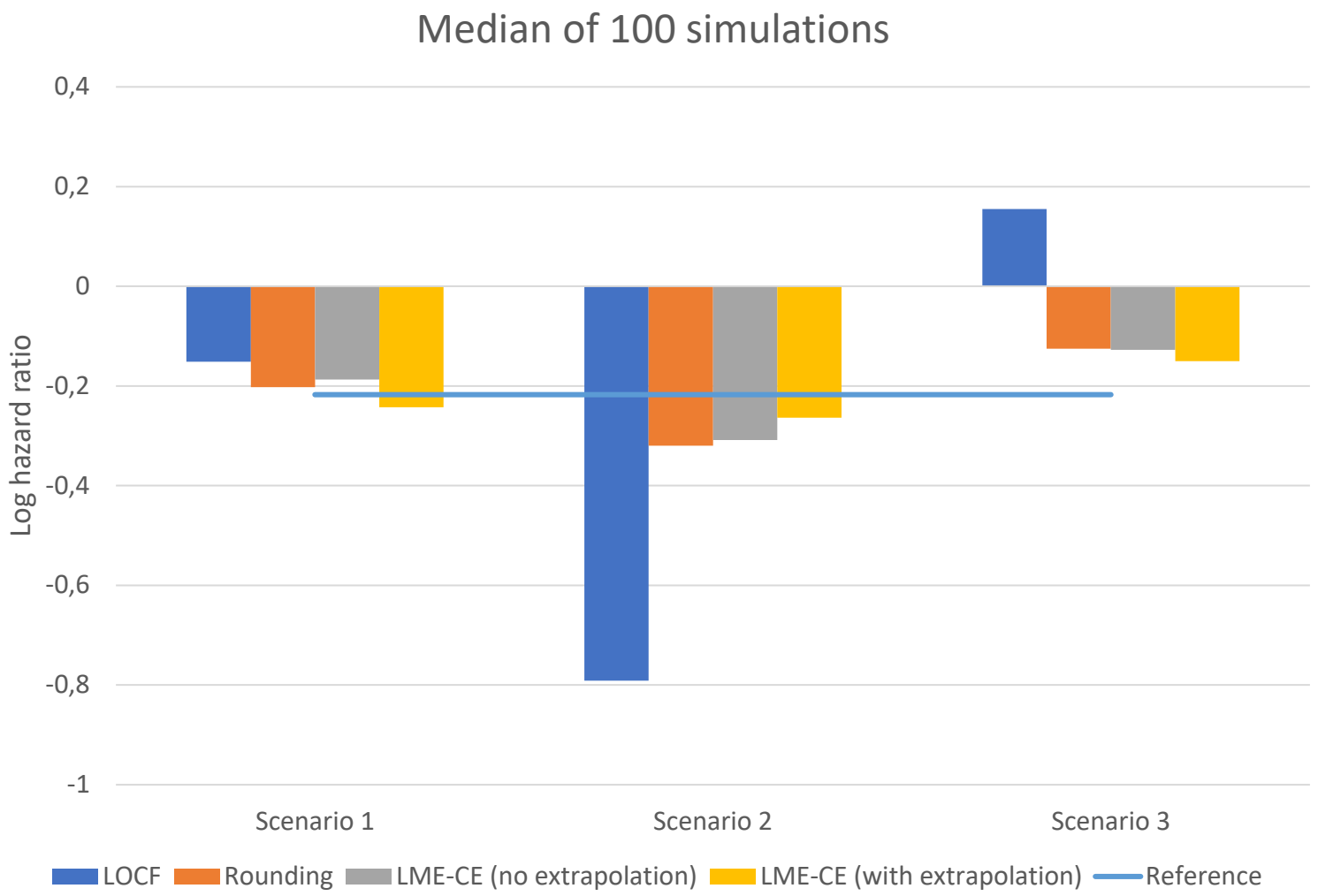
Consideration: Impute missing EDSS only between baseline and last visit, or for all visits between baseline and the maximum follow-up of 24 months?

Root Mean Squared Error of imputed EDSS



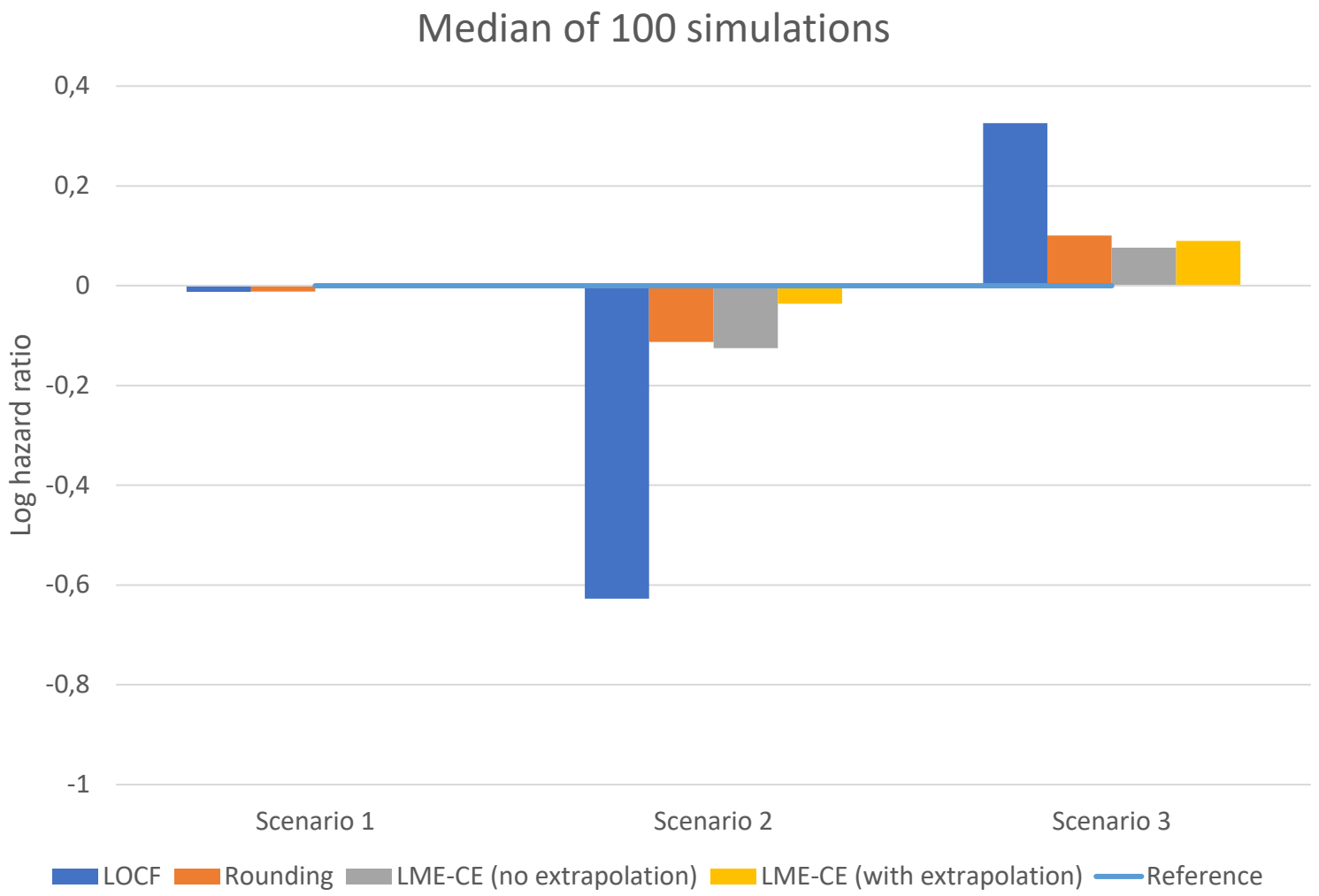
	Scenario
1	Probability of observing a visit decreases from 36% at month 1 to 6% at month 24 (~ 4 visits / 24 months)
2	**Treatment A**: probability of observing a visit at 6, 12, 18 and 24 months is 85%, 3% otherwise (~ 5 visits/24 months) **Treatment B**: probability of observing a visit at 9 and 18 months is 67%, 3% otherwise (~ 3 visits/24 months)
3	Probability of observing a visit varies as a function of age, treatment and unobserved EDSS score (~ 4 visits/24 months)

Estimated treatment effect



	Scenario
1	Probability of observing a visit decreases from 36% at month 1 to 6% at month 24 (~ 4 visits / 24 months)
2	**Treatment A**: probability of observing a visit at 6, 12, 18 and 24 months is 85%, 3% otherwise (~ 5 visits/24 months) **Treatment B**: probability of observing a visit at 9 and 18 months is 67%, 3% otherwise (~ 3 visits/24 months)
3	Probability of observing a visit varies as a function of age, treatment and unobserved EDSS score (~ 4 visits/24 months)

Estimates in the absence of Treatment effect



	Scenario
1	Probability of observing a visit decreases from 36% at month 1 to 6% at month 24 (~ 4 visits / 24 months)
2	**Treatment A**: probability of observing a visit at 6, 12, 18 and 24 months is 85%, 3% otherwise (~ 5 visits/24 months) **Treatment B**: probability of observing a visit at 9 and 18 months is 67%, 3% otherwise (~ 3 visits/24 months)
3	Probability of observing a visit varies as a function of age, treatment and unobserved EDSS score (~ 4 visits/24 months)

Final thoughts

Key findings

- LOCF is very problematic in almost all scenarios
- Rounding generally performs better than LOCF
- Multilevel modelling yields the best predictions for missing EDSS
- Better imputations do not always lead to better estimates of treatment effect
  - Need to preserve the distribution of the missing values as good as possible
  - Improvements may be dependent on the strength of prognostic factor and autocorrelation, and on the (mis)specification of the imputation model
  - Ommitting imputations beyond the last visit could lead to selection bias for patients with no follow-up visits (i.e. only baseline visit is available)

Next steps

- Evaluate the methods in real-world MS data with irregular visit patterns

Possible extensions

- Adjust for parameter uncertainty
- Evaluate coverage

Appendix

Data Generation Model

$$m_{tij} = 2.78 + \alpha_{ij} + \beta_{1j} + 0.014t - 0.007\, t\, x_{ij} + 0.018 a_{ij} + \epsilon_{tij}$$

$$\alpha_{ij} \sim \mathcal{N}(0, \sigma = 1.46)$$

$$\beta_{1j} \sim \mathcal{N}(0, \sigma = 0.20)$$

$$\boldsymbol{\epsilon_{ij}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{ij})$$

$$\boldsymbol{\Sigma}_{ij} = \begin{pmatrix} 0.5^2 & 0.8^1\, 0.5^2 & \dots & 0.8^{24}\, 0.5^2 \\ 0.8^1\, 0.5^2 & 0.5^2 & \dots & 0.8^{23}\, 0.5^2 \\ \vdots & \vdots & \ddots & \vdots \\ 0.8^{24}\, 0.5^2 & 0.8^{23}\, 0.5^2 & \dots & 0.5^2 \end{pmatrix}$$

$$a_{ij} \sim \mathcal{N}(\mu = 37.1, \sigma = 9.73)$$

$$x_{ij} \sim \text{Bernoulli}(\pi_{ij})$$

m_{tij}: EDSS score at time *t* for patient *i* in center *j*.

- For patients on DMT A ($x_{ij} = 0$), EDSS increases by 0.014 point per month, on average.
- For patients on DMT B, EDSS increases by 0.07 point per month, on average.
- Older patients have higher baseline EDSS.

Confounding: $\pi_{ij} = \text{logit}^{-1}(0.7 - 0.032 a_{ij})$