Estimation of Treatment Effects in Registry Data: Informative Patient Visits Require Careful Modeling

Debray TPA,[1] Copetti M,[2] Platt RW,[3,4] Simoneau G,[3,5] Pellegrini F,[6] de Moor C[7]

[1]University Medical Center Utrecht, Utrecht, CX, Netherlands; [2]Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, FG, Italy; [3]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada; [4]Centre for Clinical Epidemiology, Jewish General Hospital, Montreal, QC, Canada; [5]Biogen Canada, Mississauga, ON, Canada; [6]Biogen International GmbH, Zug, Baar, Switzerland; [7]Biogen, Cambridge, MA, USA

OBJECTIVE

- To evaluate existing and develop new methods for comparative effectiveness of treatments using registry data when outcomes are assessed at irregular visit schedules.

CONCLUSIONS

- Our proposed mixed-effects modeling method based on multiple imputations and rounding perform consistently better than last observation carried forward.
- Additional simulation scenarios are needed to investigate the impact of:
  - Visit frequency or of more complex informative data deletion mechanisms, in which case rounding may be more heavily impacted.
  - Model misspecification of the mixed-effects model, in which case the performance of the LME may be compromised.
- Ongoing work includes an application to real-world data in multiple sclerosis.

ISPE (2020) International Society for Pharmacoepidemiology - 36th ICPE | 2020

Motivation

- In multiple sclerosis, evidence from comparative effectiveness research derived from observational data are increasingly needed.
- However, these data sources can be prone to bias, including outcome assessment biases due to informative missingness of relevant patient outcomes.
- As a motivating example, consider the definition of time to confirmed disease progression (CDP), which requires the Expanded Disability Status Scale (EDSS) to be measured at 2 follow-up visits.
- In observational data sources, applying this definition is challenging because patient visits do not follow a planned follow-up schedule.
- Existing mitigation methods (last observation carried forward [LOCF] or rounding) derive time to CDP by mapping the EDSS scores at observed, irregular visits to a regular schedule.
- However, they may create artificial lag in time to CDP, leading to erroneous conclusions on treatment effects.

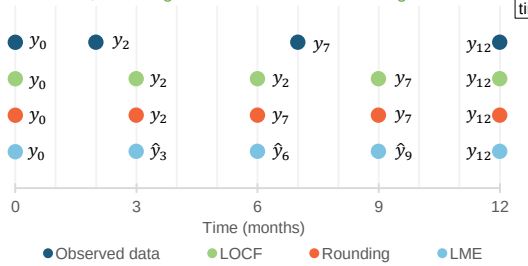
Time to Confirmed Disease Progression

It is defined as the time from baseline to an increase in:

- 1.5 points if baseline EDSS is 0
- 1.0 point if baseline EDSS is between 1.0 and 5.5
- 0.5 point if baseline EDSS is 6.0 or above

The increase must be confirmed at a visit 3 months later.

Figure 1. EDSS scores recovered at 3, 6, 9 and 12 months with LOCF, rounding and mixed-effects modeling



Proposed Mixed-Effects Modeling

- We propose to model the EDSS trajectories with a linear mixed model (LME) and generate imputations from the fitted model at the desired visit schedule.

$$y_{tij} = \alpha_{ij} + \sum_{k=1}^{K} \beta_{kj} f_k(t_{tij}) + \sum_{l=1}^{L} \delta_{lj} f_l(t_{tij}, x_{ij}) + \sum_{m=1}^{M} \gamma_m z_{mij} + \epsilon_{tij}$$

EDSS patient i, time t, center j

Patient-specific intercept $\alpha_{ij} \sim N(\mu_\alpha, \tau_\alpha^2)$

Center-specific random effects of time (t_{tij}) and treatment (x_{tij})
$\beta_{kj} \sim N(\mu_{\beta,k}, \tau_{\beta,k}^2)$ $\delta_{lj} \sim N(\mu_{\delta,l}, \tau_{\delta,l}^2)$

Effect of m prognosis factor

Error term $\epsilon_{tij} \sim N(0, \Sigma_{ij})$

- An Exponential spatial correlation structure on the errors can be used to account for within-patient correlations over time.
- Single imputations are generated as the expected value of EDSS at time t; with multiple imputations, an error term sampled from a multivariate normal distribution is added.

Simulation Study

Generate data with irregular visit patterns

- 500 patients per center, 20 centers, treatment A or B assigned as a function of age.
- EDSS generated every month for 24 months from a linear mixed model with an AR1 structure ($\rho = 0.8$).
- At each month, the generated EDSS are rounded to the nearest 0.5 and truncated between 0 and 9.5.
- Visits are deleted as a function of center, treatment, time and/or true (unobserved) EDSS scores.

1	Probability of observing a visit decreases from 38% 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)

EDSS imputations and time to CDP

- Compare LOCF, rounding and our LME with single imputation of the expected value (SI) or 20 multiple imputations (MI).
- Time to CDP derived with a 3-month confirmation window.

Estimation of treatment effect

- Time to CDP modeled with a Cox regression stratified by center, adjusted for age and baseline EDSS.
- For multiple imputations, Cox regression fitted separately in each dataset and estimates combined with Rubin's rule.

Results

Derived time to CDP (Figure 3)

- The survival curves for time to CDP derived from LME with multiple imputations or from rounding are the closest to that based on the original data.

Estimated treatment effect (Figure 4)

- LME based on multiple imputations and rounding leads to consistently less biased log-hazard ratio estimators.
- LOCF only performs well in scenario 1 when the visit mechanism does not depend on treatment or on the unobserved EDSS scores.
- LME based on single imputation performs badly, probably because the imputations do not include error terms.

Figure 2. Example EDSS trajectories in the original data (line) and EDSS at observed visits (■) in 3 scenarios.

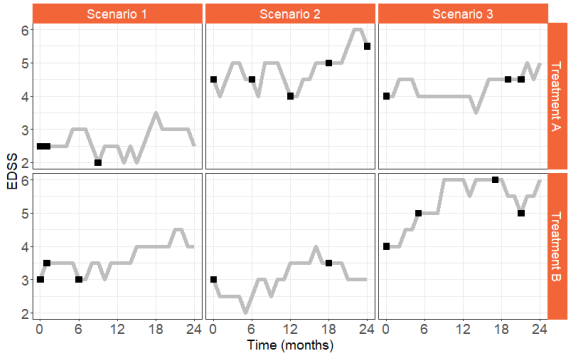


Figure 3. Survival curves for time to CDP calculated from the original data and from imputed EDSS with LOCF, rounding and LME (SI and MI).

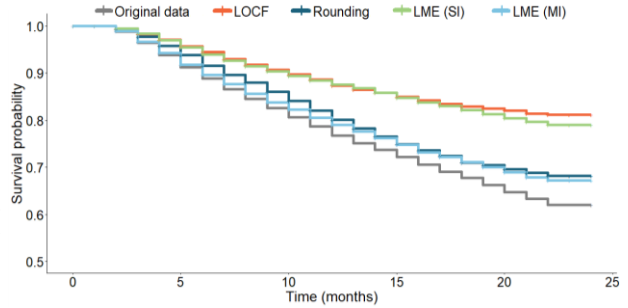
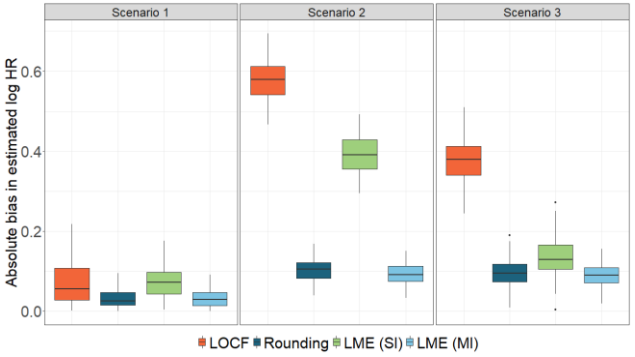


Figure 4. Absolute bias in estimated hazard ratio across 100 simulations.



Disclosures TPAD,MC,RWP: honoraria from Biogen; FP,GS,CdM: owns stocks of Biogen. Acknowledgments This study was sponsored by Biogen (Cambridge, MA, USA). Editorial support for the preparation of this poster was provided by Excel Scientific Solutions (Fairfield, CT, USA): funding was provided by Biogen.