Implementing a full random effects model (FREM) workflow in R using the *saemix* package

Alexandra Lavalley-Morelle, E. Niclas Jonsson
Pharmetheus AB, Sweden

Scan to download
E-poster

Objective

Implement a full random effect workflow within the open-source *saemix* package and evaluate its performance against standard Fixed Effects Modeling (FEM) under varying degrees of collinearity in a Repeated Time-to-Event (RTTE) modeling framework.

Conclusions

FREM methodology was successfully implemented within the saemix package and effectively handled covariate collinearity and missingness in our RTTE example. Future work will focus on a comprehensive methodological evaluation to formalize the theoretical properties of the SAEM-FREM combination and further investigate the estimation of the residual error.

Background

Covariate analysis continues to be a complex aspect of pharmacometrics modeling. The Full Random Effects Model (FREM)[1] offers a reliable approach, addressing collinearity and missing data through joint distribution estimation.

While FREM is well-established and widely used in NONMEM, the methodology is currently unavailable in R.

Data and methods

- **FREM implementation:** Used the *saemix*[2] multi-response extension[3] (SAEM algorithm[4]) and treated covariates as additional dependent variables to estimate a joint multivariate normal distribution rather than explicit fixed effects.
- **Simulation design:** Generated 100 datasets (500 subjects each) featuring a Repeated Time-to-Event (RTTE) outcome with a constant baseline hazard (h_0) and inter-individual variability: $h_{0i} = h_0 \times \exp(\eta_i)\,,\ \ \eta_i \sim N(0, \omega)$
- **Covariate structure:** Simulated 20 covariates grouped into four correlation blocks (Figure 1). The true underlying model was driven by four active covariates (one from each correlation block):

$$h_i(t) = h_{0_i} \times \exp(\beta_1 Cov_1 + \beta_6 Cov6 + \beta_{11} Cov11 + \beta_{16} Cov16)$$

- Missingness scenarios: Three data quality levels were simulated on true active covariates: High (0%) Medium (20%) Low (50%)

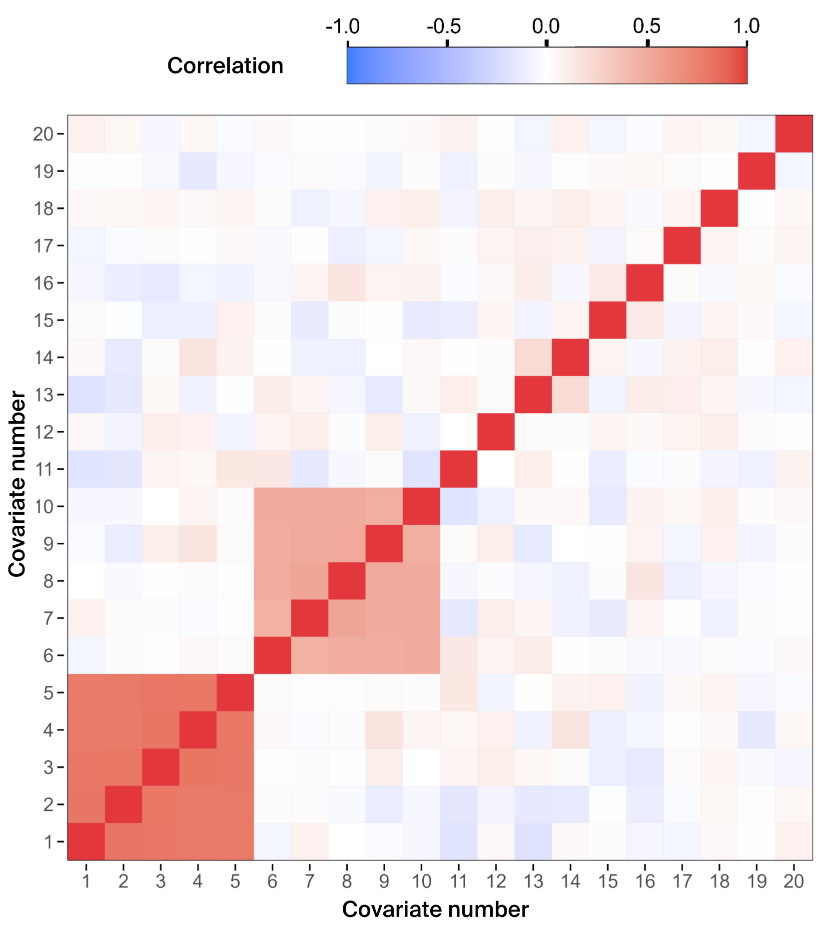

Figure 1. Correlation structure between simulated covariates.

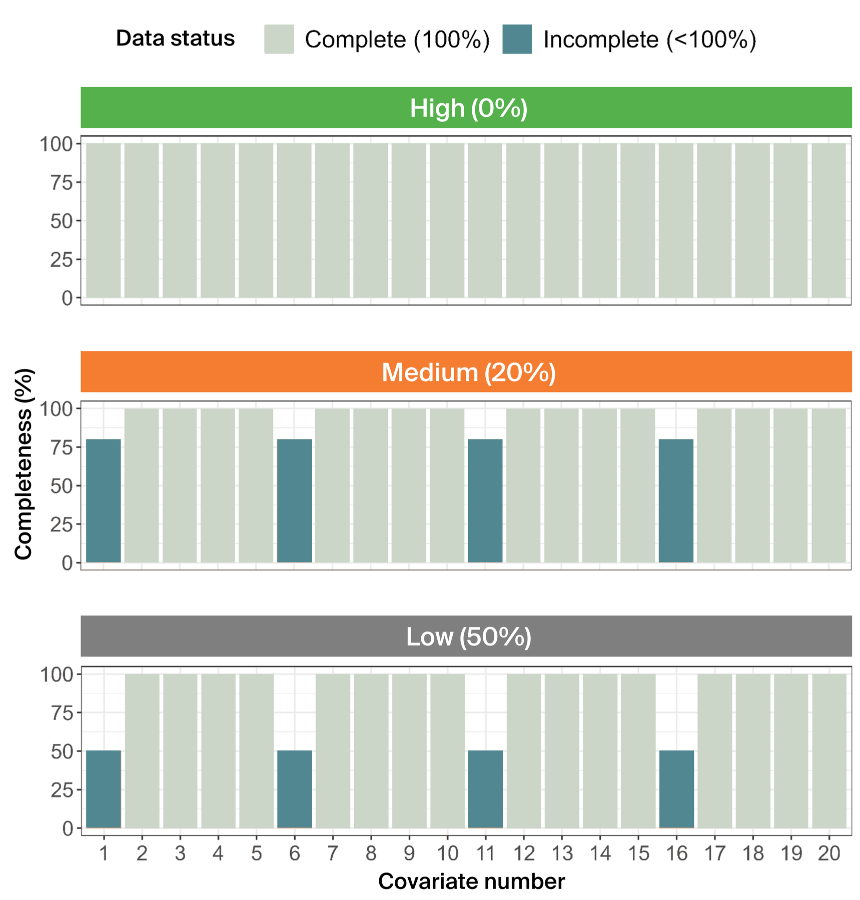

Figure 2. Covariate completeness profiles across scenarios.

- **Model comparison (FEM vs. FREM):** The impact of collinearity and missingness was evaluated by comparing standard Fixed Effects Modeling (FEM) with FREM, incrementally adding 1 to 5 correlated covariates per block across all data quality scenarios.
- **Performance metrics:** Quantified bias and precision using Relative Mean Error (RME) and Relative Root Mean Squared Error (RRMSE), comparing the estimated linear predictor ($\widehat{LP}$) to the true predictor (LP), for a typical subject:

$$LP = X \times \beta^t \text{ and } \widehat{LP} = X \times \hat{\beta}^t \text{ where } X = [\overline{Cov_1}, \ldots, \overline{Cov_n}]$$

Results

FREM models were successfully implemented, treating covariates as additional dependent variables alongside the RTTE outcome and estimating the joint multivariate normal distribution. The covariate effects were derived from the full variance-covariance matrix between the random effects and the covariates.

Impact on

- **Collinearity** — FREM maintains near-zero bias regardless of the number of correlated covariates included. In contrast, standard FEM exhibits progressively larger bias as collinearity increases.
- **Missingness** — The traditional FEM approach breaks down under missing data (RME = 34.3% in the worst-case scenario involving 50% missing data on the four active covariates and 3 correlated covariates per correlation block), while the FREM approach maintains robust and unbiased estimations even at 50% missingness (RME = 1.2% in the same scenario).
- **Accuracy** — Across both medium and low-quality data, FREM demonstrates consistently superior precision (lower RRMSE) compared to standard fixed-effects modeling.

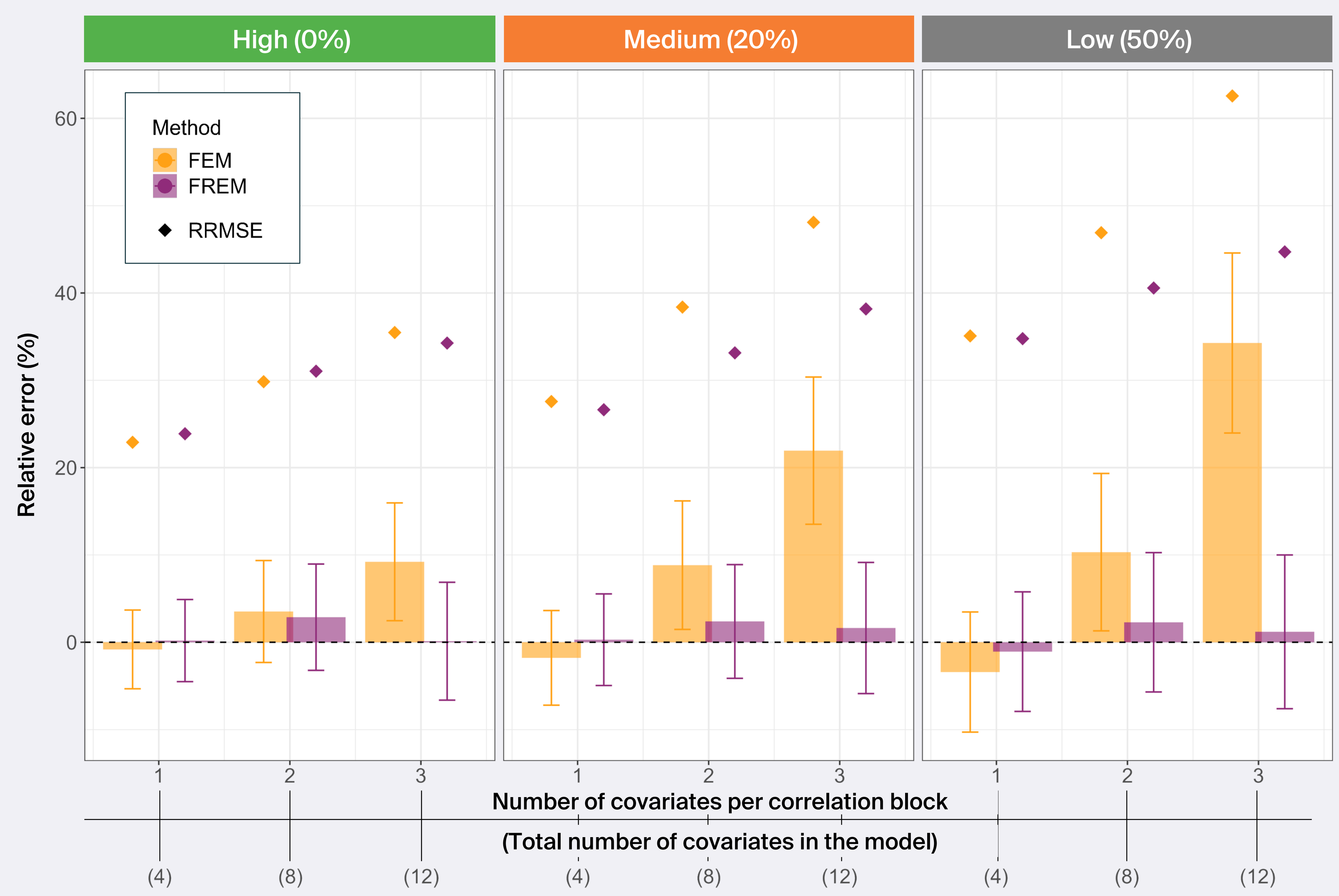

Figure 3. Relative mean error (RME) and relative root mean square error (RRMSE) on the linear predictor depending on the various scenarios, obtained across 100 simulations. The error bar show the 95% confidence interval around the RME.

Implementation challenges

- Models including 4 or 5 covariates per correlation blocks (i.e. 16 or 20 covariates in total) had issues related to convergence stability, possibly caused by over-parameterization.
- Traditional FREM implementations (using IMP MAP algorithm) recommend covariate residual error variances to be fixed at negligible values (e.g. 10^{-6}).
  - → This process created significant bias with the SAEM algorithm, depending on the proportion of missing data in the covariates.
  - → We adapted the process to natively estimate these variances, eliminating the bias regardless of missingness.
- Software-independent caution is advised for SAEM / FREM combinations

References

1. Nyberg et al. (2024), doi:10.1002/sim.9979
2. Comets et al. (2017), doi:10.18637/jss.v080.i03
3. Lavalley-Morelle et al. (2024), doi:10.1016/j.cmpb.2024.108112
4. Delyon et al. (1999), doi:10.1214/aos/1018031103
5. Jonsson et al. (2024), doi:10.1002/psp4.13110

Contact

alexandra.lavalley-morelle@pharmetheus.com
www.pharmetheus.com

Pharmetheus

Supplementary material

Implementing a full random effects model (FREM) workflow in R using the *saemix* package

Additional results for all covariates scenarios

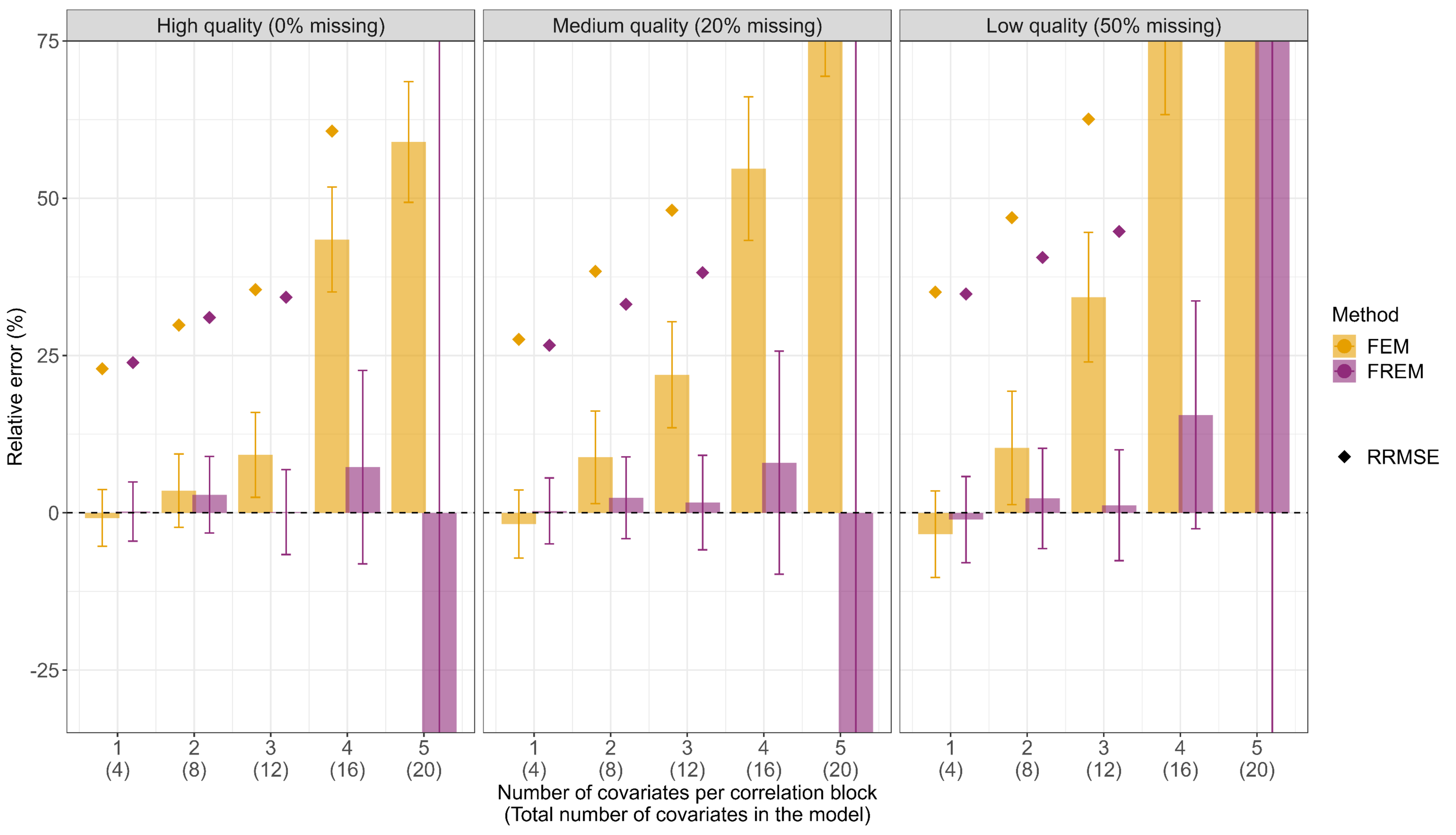

Figure S3. Relative mean error (RME) and relative root mean square error (RRMSE) on the linear predictor depending on all the scenarios, obtained across 100 simulations. The error bar show the 95% confidence interval around the RME.

- High, though statistically non-significant, bias in high-dimensional covariate spaces (4–5 covariates per correlation block, meaning Ω of size 17x17 or 21x21).

Investigations on σ for the covariate models

- **Covariate models:** The model for covariate x can be defined as:

$$Cov_{i,x} = \mu_x + \eta_{i,x} + \sigma_x$$

- **Estimation strategies:** Evaluated under the medium data quality scenario (true active covariates only):
  - Fixed σ_x : Evaluated across a grid of values ($\sigma_x \in \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
    Rationale: Fixing σ_x near zero resolves structural unidentifiability; with only a single baseline observation per individual, the model cannot distinguish between IIV and residual error.
  - Estimated σ_x : Full estimation of σ_x parameters
- **Evaluation metrics:** Relative Mean Error (RME) and Relative Root Mean Square Error (RRMSE).

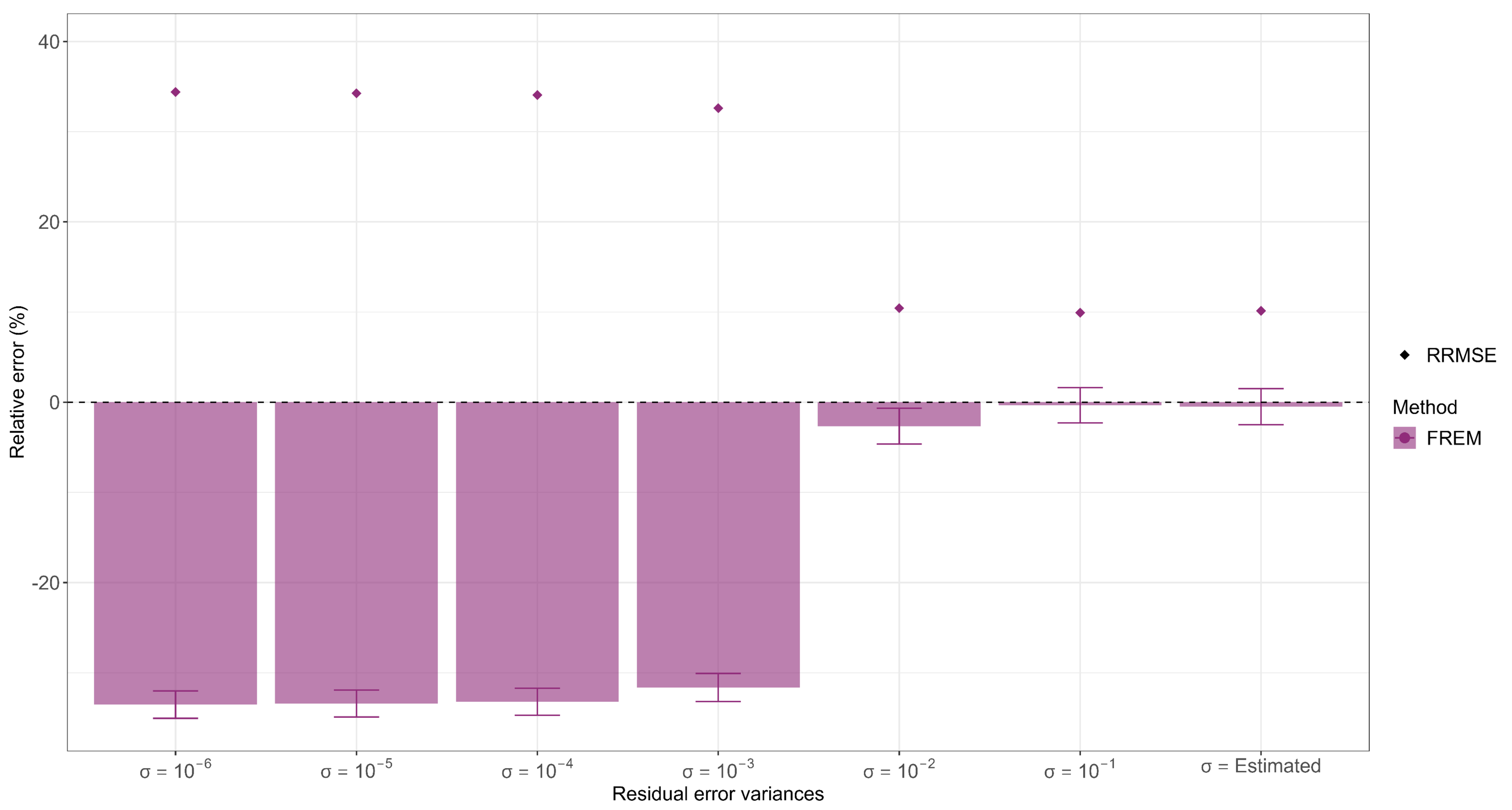

Figure S1. Relative mean error (RME) and relative root mean square error (RRMSE) on the linear predictor depending on various scenarios by fixing or estimating the residual error variances for the covariate models (σ), obtained across 100 simulations. The error bar show the 95% confidence interval around the RME.

- Low fixed σ_x values introduced a high bias in the estimation. In particular:
  - ω relative to the baseline hazard h_0 was over-estimated
  - Covariate effects (β_x) were under-estimated
- Estimated values of σ_x were close to 0.15 in mean, and avoided bias

Example of a FREM implementation using *saemix*

Long data format

ID	Tstart	Tstop	Event	Age	Weight	Sex
1	0.00000000	7.01828647	1	27.92934	80.82272	0
1	7.01828647	10.00000000	0	27.92934	80.82272	0
2	0.00000000	8.77340982	1	42.77429	83.36123	1
2	8.77340982	10.00000000	0	42.77429	83.36123	1

ID	Tstart	Tstop	Value	ytype
1	0.0000000	0.0000000	27.92934	1
1	0.0000000	0.0000000	80.82272	2
1	0.0000000	0.0000000	0.00000	3
1	0.0000000	7.0182865	1.00000	4
1	7.0182865	10.0000000	0.00000	4
2	0.0000000	0.0000000	42.77429	1
2	0.0000000	0.0000000	83.36123	2
2	0.0000000	0.0000000	1.00000	3
2	0.0000000	8.7734098	1.00000	4
2	8.7734098	10.0000000	0.00000	4

Model implementation

Structural model definition

```
rtte_model_frem <- function(psi, id, x) {
  ytype <- x[,"ytype"]
  tstart <- x[,1][ytype==4]
  tstop <- x[,2][ytype==4]
  event <- x[,3] [ytype==4]

  cens<-which(event==0) # censored ones
  ind <- which(event==1) # event ones

  lambda_i <- psi[id, 1][ytype==4]
  mean_age  <- psi[id, 2]
  mean_weight  <- psi[id, 3]
  mean_sex  <- psi[id, 4]

  haz <- exp(lambda_i)
  Hstart <- haz*tstart # cumulative hazard up to tstart
  Hstop <- haz*tstop # cumulative hazard up to tstop

  logpdf <- rep(0,length(tstart))
  # likelihood contributions for censored observations
  logpdf[cens] <- -Hstop[cens] + Hstart[cens]
  # likelihood contributions for event observations
  logpdf[ind] <- -Hstop[ind] + Hstart[ind] + log(haz[ind])

  ypred <- rep(0,nrow(x))
  ypred[ytype == 4] = logpdf
  ypred[ytype == 1] <- mean_age[ytype == 1]
  ypred[ytype == 2] <- mean_weight[ytype == 2]
  ypred[ytype == 3] <- mean_sex[ytype == 3]

  return(ypred)
}
```

saemix specification

```
# Initial values: [Intercept, Slope, Mean_Age, Mean_Sex]
param_frem <- c(0.5, mean(df_frem_rtte$Value[df_frem_rtte$ytype==1]),
                mean(df_frem_rtte$Value[df_frem_rtte$ytype==2]),
                mean(df_frem_rtte$Value[df_frem_rtte$ytype==3]))

# FULL Covariance matrix (1s everywhere)
# This tells SAEMIX to estimate the correlation between parameters and covariates
m_frem <- matrix(1, ncol = 4, nrow = 4)

model_saemix_frem <- saemixModel(
  model = rtte_model_frem,
  description = "RTTE FREM model",
  modeltype = c("structural","structural","structural","likelihood"),
  psi0 = matrix(param_frem, ncol = 4, byrow = TRUE,
                dimnames = list(NULL, c("lambda", "mAge", "mWeight", "mSex"))),
  transform.par = c(0, 0, 0, 0), # 0: Normal, 1: Log-normal
  covariance.model = m_frem, # Full matrix for FREM
  fixed.estim = c(1, 0, 0, 0), # 0: Fixed, 1: Estimated
  error.model = c("constant","constant","constant"),
  error.init = c(0.01,0,0.01,0,0.01,0,0,0),
  fixed.res = c(1,0,1,0,1,0,0,0), # 0: Fixed, 1: Estimated
)

saemix.optionsNOFIM<-saemixControl(seed=12345, map=F, fim=F,
                                   ll.is=F, save.graphs = F,
                                   nb.chains = 3, nbiter.saemix = c(500,150))
fit_FREM <- saemix.multi(model = model_saemix_frem,
                         data = rtte_saemix_frem,
                         control = saemix.optionsNOFIM)
```

Investigations on the missingness proportion

- **Estimation strategies**: Evaluated under the scenario involving only the true active covariates, and fixing all the σ_x to 10^{-6} :
  - Varying the proportion of missing data in the covariates (from 5% to 50%, by 5%)
- **Evaluation metrics**: Relative Mean Error (RME) and Relative Root Mean Square Error (RRMSE).

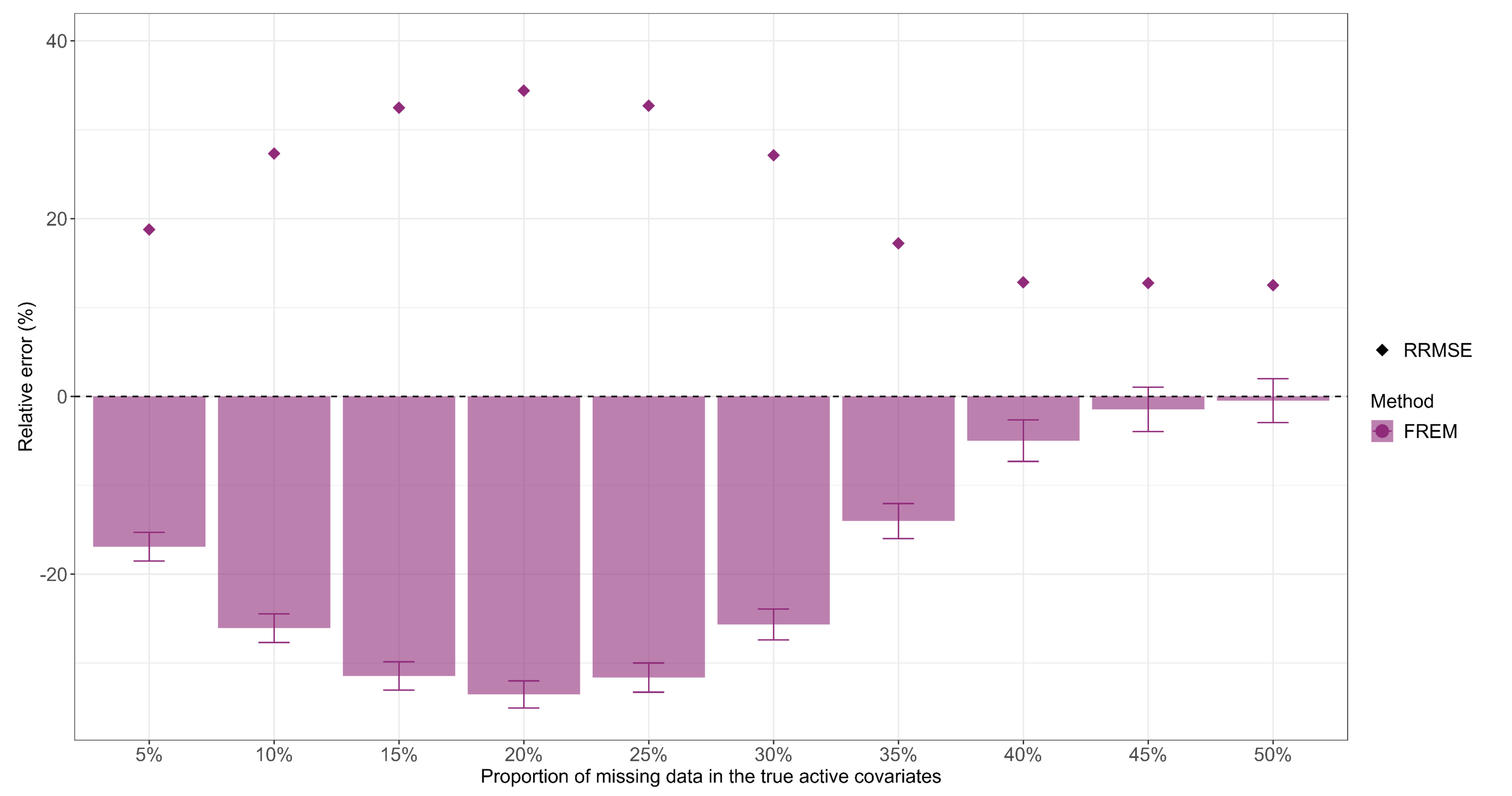

Figure S2. Relative mean error (RME) and relative root mean square error (RRMSE) on the linear predictor depending on various scenarios of missingness, obtained across 100 simulations. The error bar show the 95% confidence interval around the RME.

- Bias increased proportionally with the missing data rate up to 40%, after which a distinct/paradoxical drop in bias was observed.