Wandering the Bayesian countryside with ‘brms’

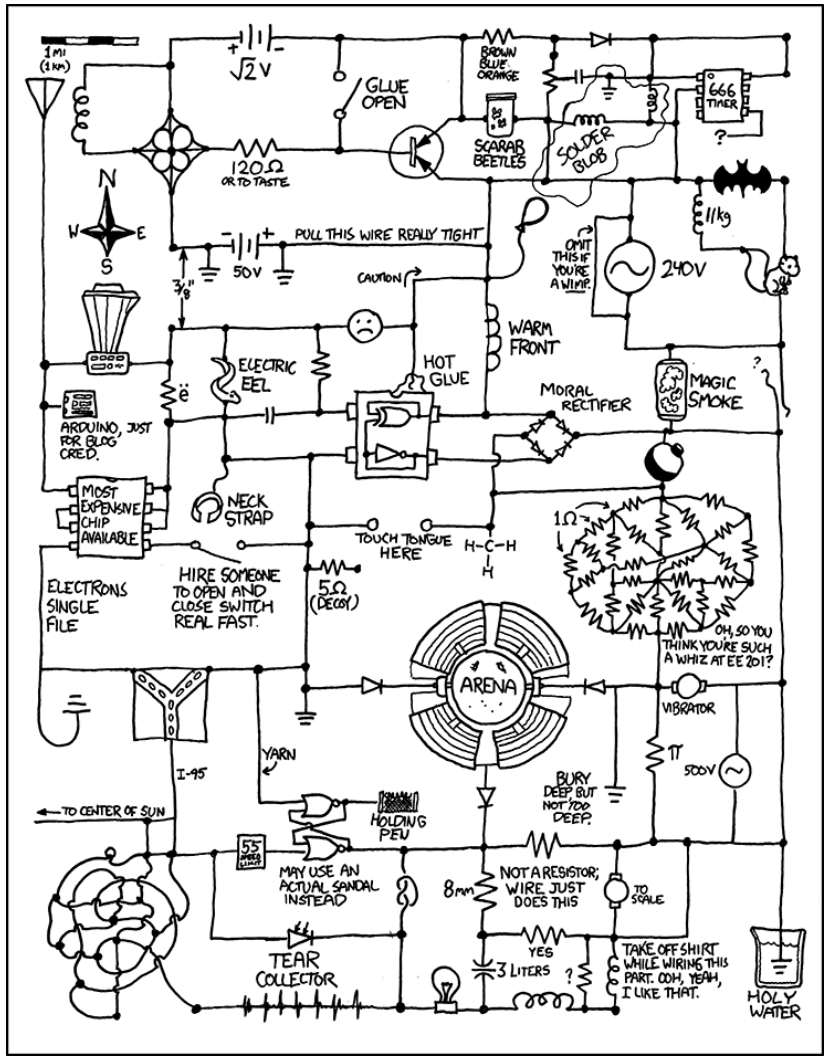
- Example of a nonlinear mixed-effect model in clinical oncology -

Francois Mercier, Francesco Brizzi

May 22, 2019 – Lyon

Acknowledgments

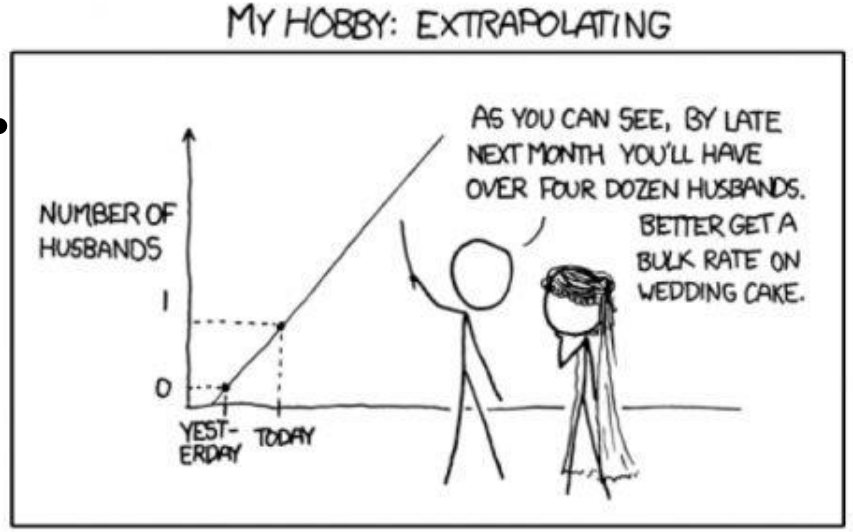
- Stan dev. team (including also support team for Wiki and Discourse)
- Paul Bürkner, author/maintainer of ‘brms’
- Blogs from Andrew Gelman, Markus Gesmann (‘Mages’), Solomon Kurz, Matti Vuorre, Kristoffer Magnusson (‘Rpsychologist’), Henrik Singmann, Daniel Lüdecke (‘Strenge Jacke’), … and many others
- Project Data Sphere
- R and Rstudio core teams



Mechanistic models
QSP
Systems Biology

Nonlinear models to describe and/or understand disease progression

Phenomenological or empirical models
Linear regression

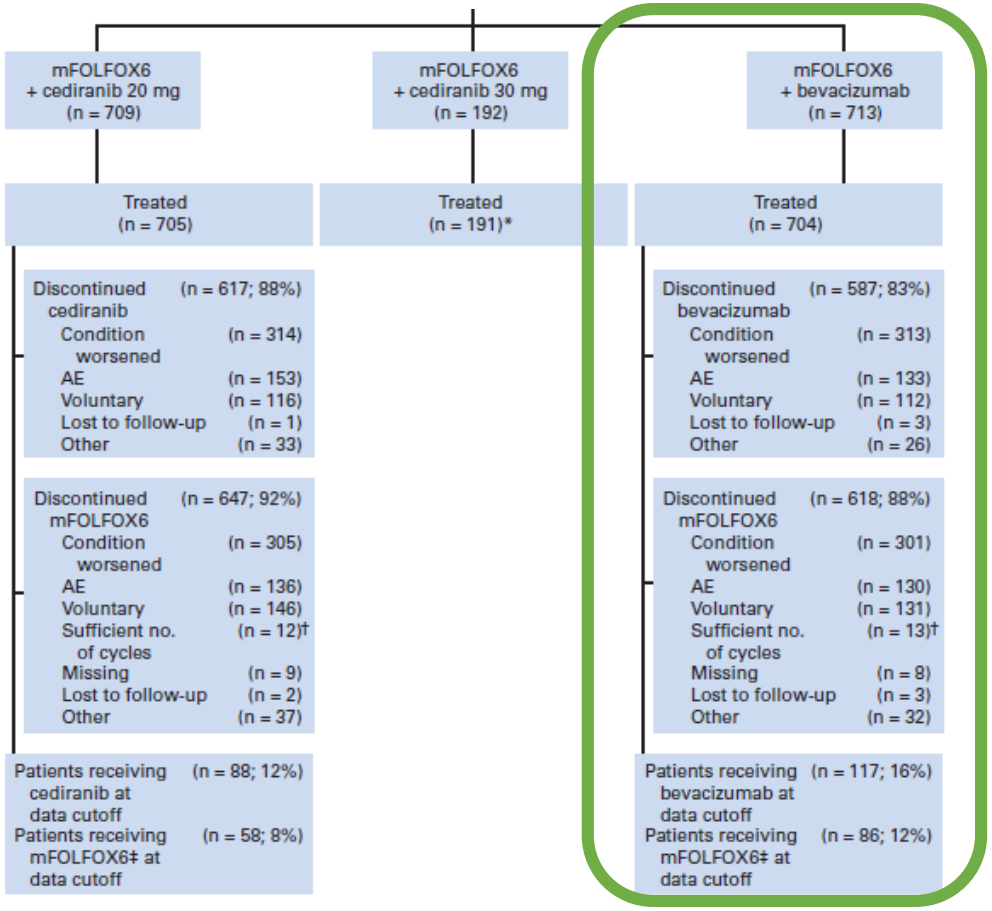


Motivating example

mCRC patients treated with bevacizumab+chemo

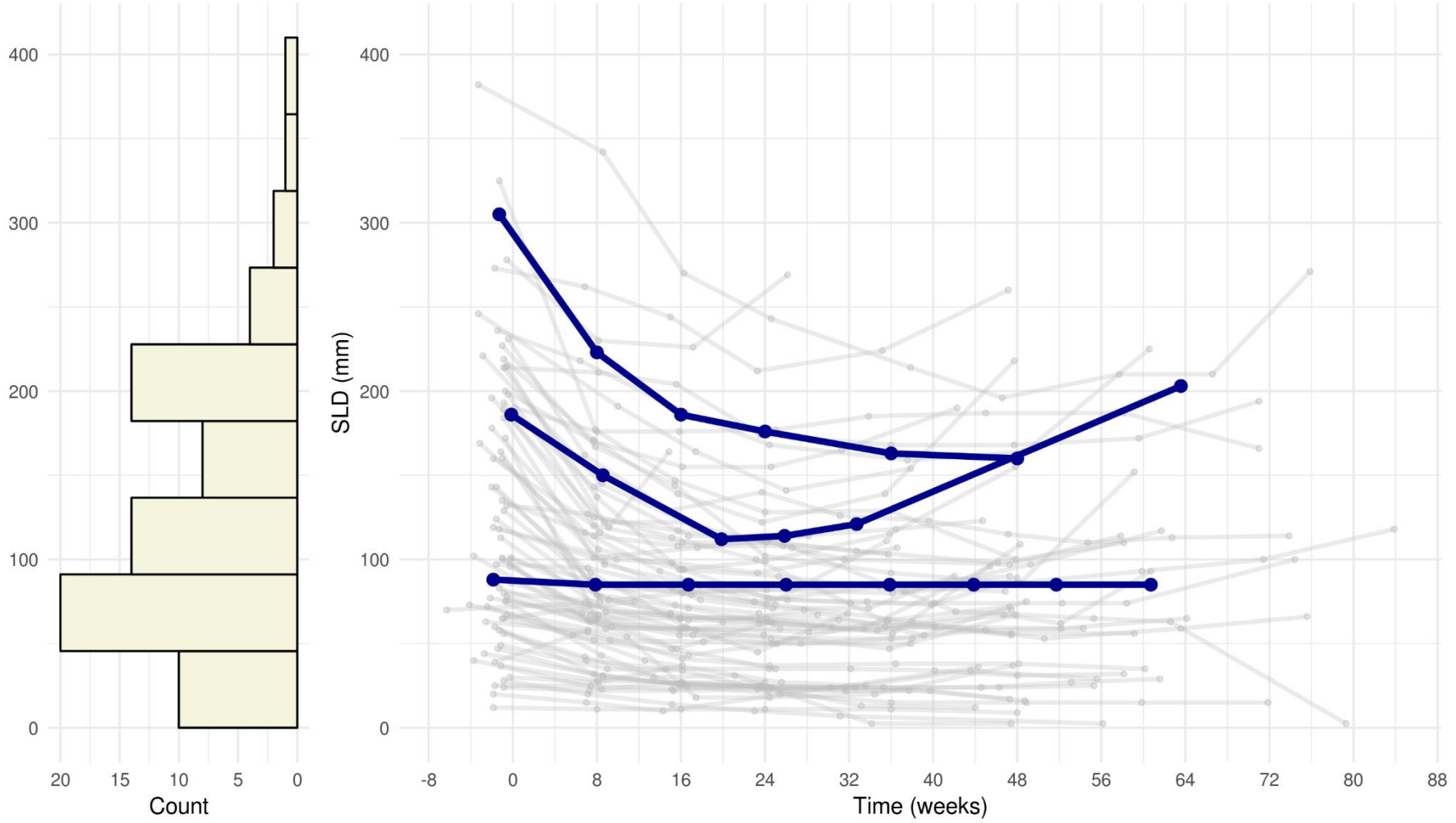
- Tumor size measured on MRI or CT scans, collected every 8 weeks until week 24 and subsequently every 12 weeks until disease progression or death
- Response variable: SLD = sum of lesion diameter (mm)
- SLD is correlated with overall survival; understanding therapeutic effect on SLD time dynamics is critical

HORIZON III study data [1]



[1] Schmoll *et al*. J Clin Oncol 2012, 30:3588-3595

Data display gives insights



[tidyverse+patchwork]

Our goals

1. **Summarize** the time trend at the **population** and at the **individual** levels

2. Investigate if/which intrinsic factors (a.k.a. covariates) may contribute to **explaining the inter-individual variability** (IIV)

Our goals ➧ Our plan

1. **Summarize** the time trend at the **population** and at the **individual** levels
    - Given a set of **candidate models**, retain the one offering the best **predictive performance**

2. Investigate if/which intrinsic factors (a.k.a. covariates) may contribute to **explaining the IIV**
    - From the best model (among candidates), extract **individual parameter estimates summarizing** the time dynamics of SLD and assess **correlation** with relevant **covariates**

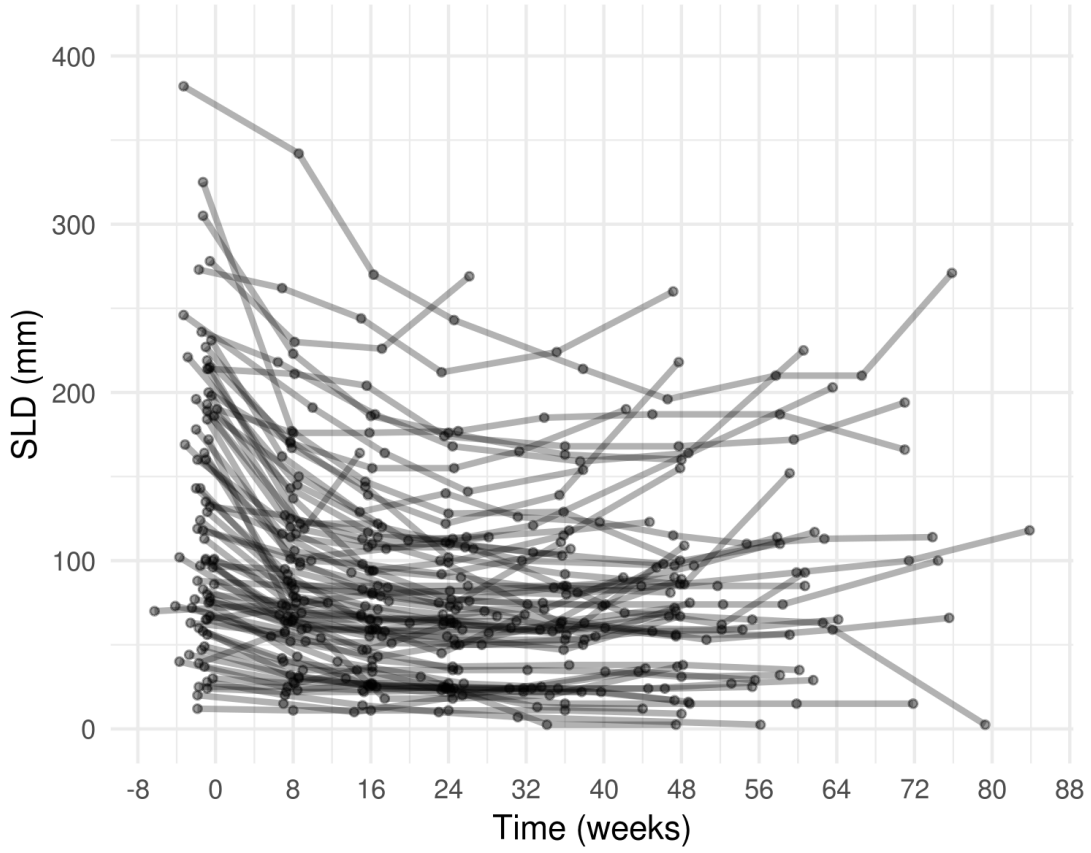
Our goals ➧ Our plan

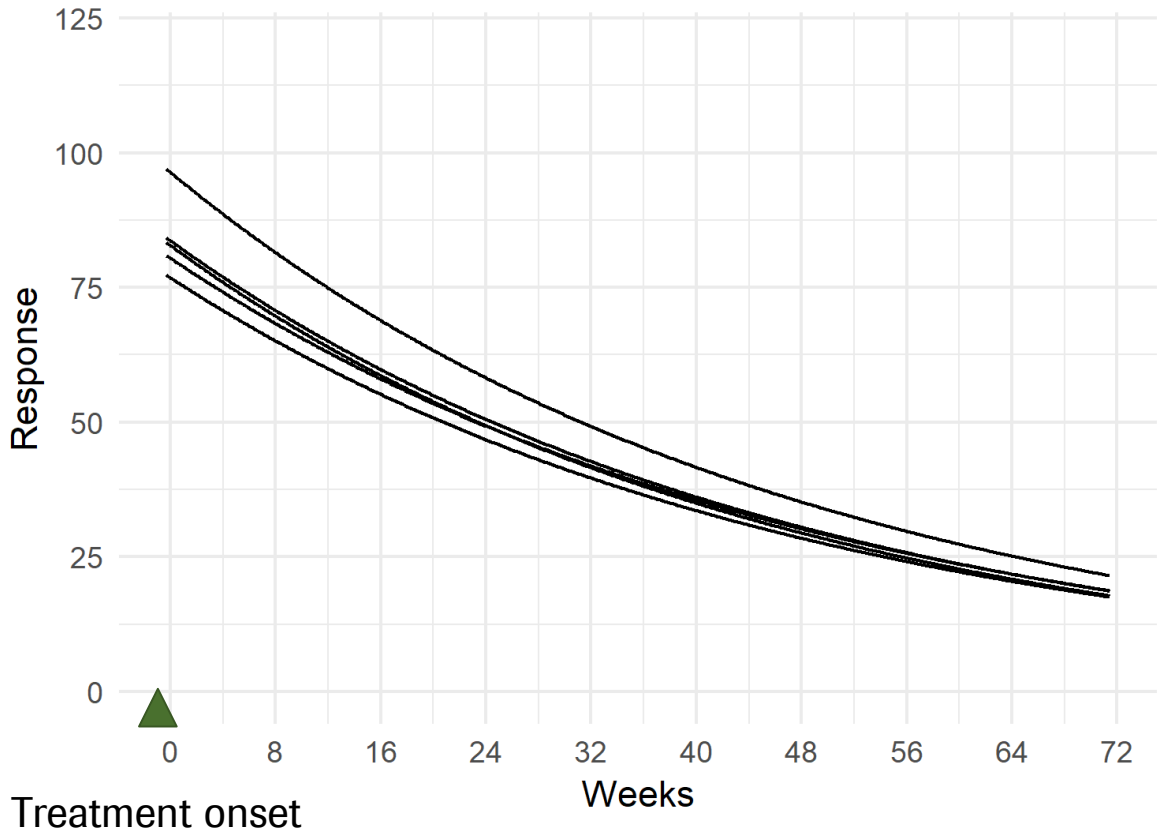
1. **Summarize** the time trend at the **population** and at the **individual** levels
    - Given a set of **candidate models**, retain the one offering the best **predictive performance**

2. Investigate if/which intrinsic factors (a.k.a. covariates) may contribute to **explaining the IIV**
    - From the best model (among candidates), extract **individual parameter estimates summarizing** the time dynamics of SLD and assess **correlation** with relevant **covariates**

Candidate models

$$SLD_{ij} = BAS_i \cdot exp\left(-KS_i \cdot t_{ij}\right) + \varepsilon_{ij}$$





Candidate models

Name	Structural form
Exponential decay	$SLD_{ij} = BAS_i \cdot exp\left(-KS_i \cdot t_{ij}\right) + \varepsilon_{ij}$
Stein-Fojo[1]	$SLD_{ij} = BAS_i \cdot \left(exp\left(KG_i \cdot t_{ij}\right) + exp\left(-KS_i \cdot t_{ij}\right) - 1\right) + \varepsilon_{ij}$
Wang[2]	$SLD_{ij} = BAS_i \cdot exp\left(KS_i \cdot t_{ij}\right) + KG_i \cdot t_{ij} + \varepsilon_{ij}$
Generalized Stein-Fojo[3]	$SLD_{ij} = BAS_i \cdot \left((1-f) \cdot exp\left(KG_i \cdot t_{ij}\right) + f \cdot exp\left(-KS_i \cdot t_{ij}\right)\right) + \varepsilon_{ij}$

BAS=SLD at time 0 (mm); KS=Shrinkage rate (1/day); KG= Growth rate (1/day); f =Fraction (responder); ε=Residual (mm).

[1] Stein *et al*. DOI: 10.1634/theoncologist.2008-0016; [2] Wang et al. DOI: 10.1038/clpt.2009.64; [3] Chaterjee *et al*. DOI:10.1002/psp4.12140.

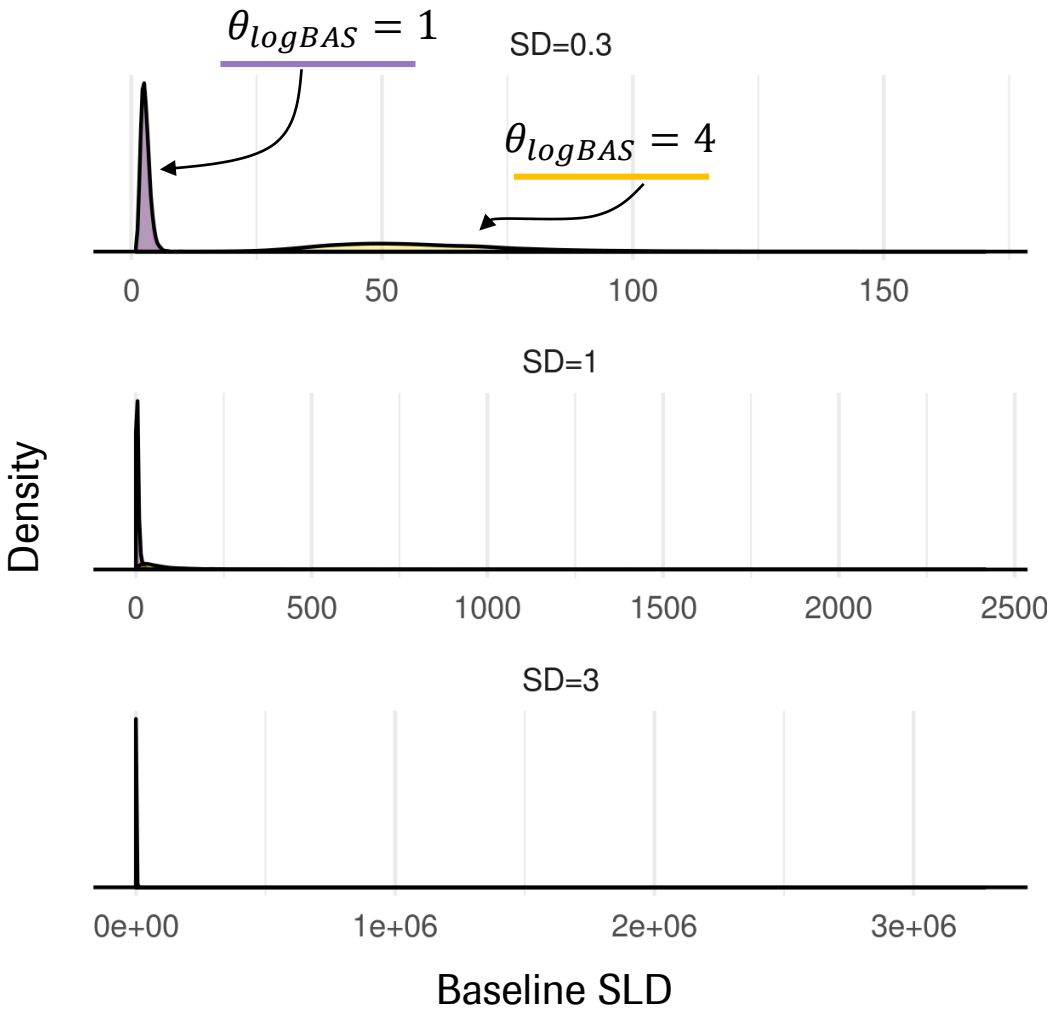
Candidate models

Name	Structural model
Generalized Stein-Fojo	$SLD_{ij} = BAS_i \cdot \left((1-f)\cdot exp\left(KG_i \cdot t_{ij}\right) + f \cdot exp\left(-KS_i \cdot t_{ij}\right)\right) + \varepsilon_{ij}$

Transformation	Param. P	Distribution of P	Distribution of θ	Distribution of ω
$BAS_i = exp(logBAS_i)$	$logBAS$	$\mathbb{N}\left(\theta_{logBAS}, \omega^2_{logBAS}\right)$	$\mathbb{N}\left(T_{logBAS}, {S_{logBAS}}^2\right)$	$\mathbb{N}\left(V_{logBAS}, {U_{logBAS}}^2\right)^+$
$KS_i = exp(logKS_i)$	$logKS$	$\mathbb{N}\left(\theta_{logKS}, \omega^2_{logKS}\right)$	$\mathbb{N}\left(T_{logKS}, {S_{logKS}}^2\right)$	$\mathbb{N}\left(V_{logKS}, {U_{logKS}}^2\right)^+$
$KG_i = exp(logKG_i)$	$logKG$	$\mathbb{N}\left(\theta_{logKG}, \omega^2_{logKG}\right)$	$\mathbb{N}\left(T_{logKG}, {S_{logKG}}^2\right)$	$\mathbb{N}\left(V_{logKS}, {U_{logKS}}^2\right)^+$
$f_i = invlogit(logitf_i)$	$logitf$	$\mathbb{N}\left(\theta_{logitf}, \omega^2_{logitf}\right)$	$\mathbb{N}\left(T_{logitf}, {S_{logitf}}^2\right)$	$\mathbb{N}\left(V_{logitf}, {U_{logitf}}^2\right)^+$
-	ε	$\mathbb{N}(0, \sigma^2)$	-	-

Note: sample from a MVN distribution to account for the correlation between hyper-parameters

Beware of the transformation!



Example:

Baseline SLD (BAS) are in a [2, 400] mm range, with mode typically between 50 and 100 mm.

↳ $logBAS$ in [log(2), log(400)]≈[0.7, 6],

↳ Mode: [log(50), log(100)]≈[4, 4.6]

As $\mathbb{N}(0,1)$ ranges approx. from -3 to +3

↳ $logBAS \sim 4 + \mathbb{N}(0,1)$ approx. ranges in [1, 7]

↳ Equivalent to BAS covering the range [2.7, 1096]

To become familiar with distributions: https://statdist.ksmzn.com/ and https://github.com/jhelvy/stanTuner

Fitting the exponential decay model with ‘brms’

```
mExpDecay<-bf(SLD~exp(lbas)*exp(-(exp(lks)/1000)*TIME),
              lbas~1+(1|UID), lks~1+(1|UID), nl=TRUE)
```

Structural+error models

```
Priorm01<-c(set_prior("normal(4, 1)", class="b", nlpar="lbas"),
            set_prior("normal(0, 1)", class="b", nlpar="lks"),
            set_prior("normal(0, 1)", class="sd", nlpar="lbas"),
            set_prior("normal(0, 1)", class="sd", nlpar="lks"))
```

Priors

```
m01<-brm(mExpDecay, data=trndf, prior=Prior01, family=gaussian(),
       iter=2000, chains=4, warmup=1000, seed=1234, cores=4)
```

Run

[brms]

Checking the convergence

[shinystan]

Model building

#ID	Structure	Family	Covariance	LOO-CV IC	SE	10x-CV IC	SE
00	Exp.decay	Gaussian	none	4095	61		
01	Exp.decay	Student	none	3840	60		
02	Stein-Fojo	Student	none	3469	55	4863	38
03	Stein-Fojo	Student	r(lbas, lks, lkg)	3473	54		
04	Wang	Student	none	4195	54		
05	Wang	Student	r(lbas, lks, lkg)	4197	54		
07	Generalized Stein-Fojo	Student	none	3431	56	4849	38
08	Generalized Stein-Fojo	Student	r(lbas, lks, lkg)	3417	57	4841	36

Note: Run time for 4 chains, 8000 iterations, on my average laptop (4 cores in parallel) ranges from 1.5 (model #01) to 1h25 min (model #08).

[brms+loo] https://avehtari.github.io/modelselection/

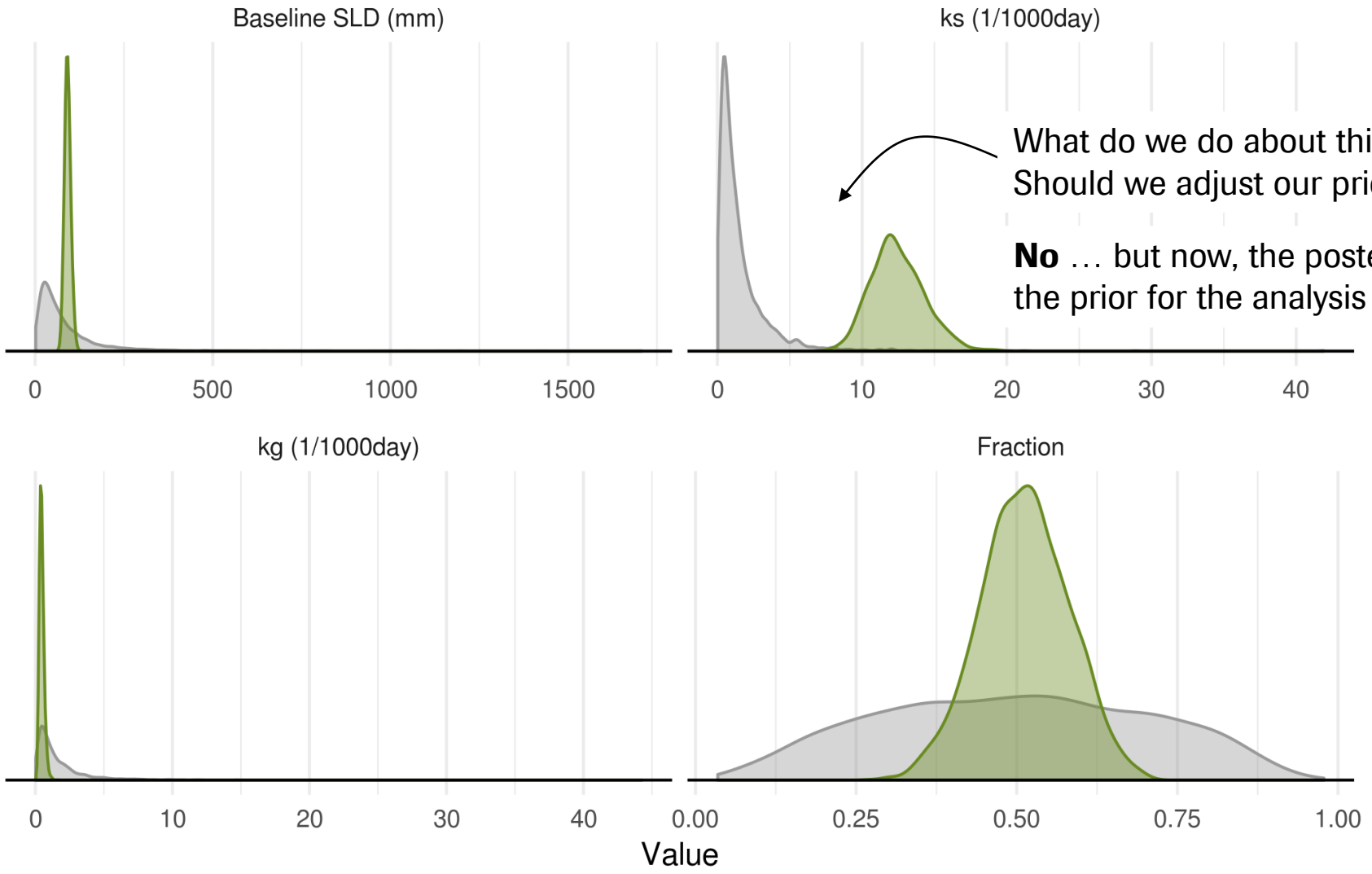
Model estimates summary

Generalized Stein-Fojo model (model #08)

```
 Family: student
  Links: mu = identity; sigma = identity; nu = identity
Formula: SLD2 ~ exp(lbas) * ((1 - (1/(1 + exp(-logitf)))) * exp((exp(lkg)/1000) * TIME) + (1/(1 + exp(-logitf))) * exp(-(exp(lks)/1000) * TIME))
         lbas ~ 1 + (1 | ID1 | UID)
         lks ~ 1 + (1 | ID1 | UID)
         lkg ~ 1 + (1 | ID1 | UID)
         logitf ~ 1 + (1 | ID1 | UID)
   Data: trndf (Number of observations: 460)
Samples: 4 chains, each with iter = 5000; warmup = 3000; thin = 1;
         total post-warmup samples = 8000

Group-Level Effects:
~UID (Number of levels: 75)
                                  Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
sd(lbas_Intercept)                    0.79      0.07     0.67     0.95        280 1.02
sd(lks_Intercept)                     0.74      0.14     0.48     1.03        250 1.02
sd(lkg_Intercept)                     1.75      0.28     1.25     2.35        353 1.01
sd(logitf_Intercept)                  1.92      0.30     1.40     2.56        253 1.02
cor(lbas_Intercept,lks_Intercept)    -0.25      0.16    -0.53     0.09        650 1.00
cor(lbas_Intercept,lkg_Intercept)     0.27      0.14    -0.02     0.53       1193 1.00
cor(lks_Intercept,lkg_Intercept)     -0.42      0.22    -0.75     0.11        194 1.02
cor(lbas_Intercept,logitf_Intercept)  0.05      0.12    -0.21     0.28       1033 1.00
cor(lks_Intercept,logitf_Intercept)  -0.53      0.20    -0.82    -0.07        137 1.03
cor(lkg_Intercept,logitf_Intercept)   0.72      0.11     0.46     0.88        503 1.01

Population-Level Effects:
                Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
lbas_Intercept      4.51      0.09     4.33     4.68        158 1.01
lks_Intercept       2.51      0.15     2.21     2.79        339 1.01
lkg_Intercept      -0.84      0.36    -1.61    -0.19        559 1.01
logitf_Intercept    0.05      0.28    -0.47     0.65        562 1.01

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
sigma     1.47      0.20     1.13     1.93        412 1.01
nu        1.05      0.05     1.00     1.18       2884 1.00

Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample
is a crude measure of effective sample size, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

[brms]

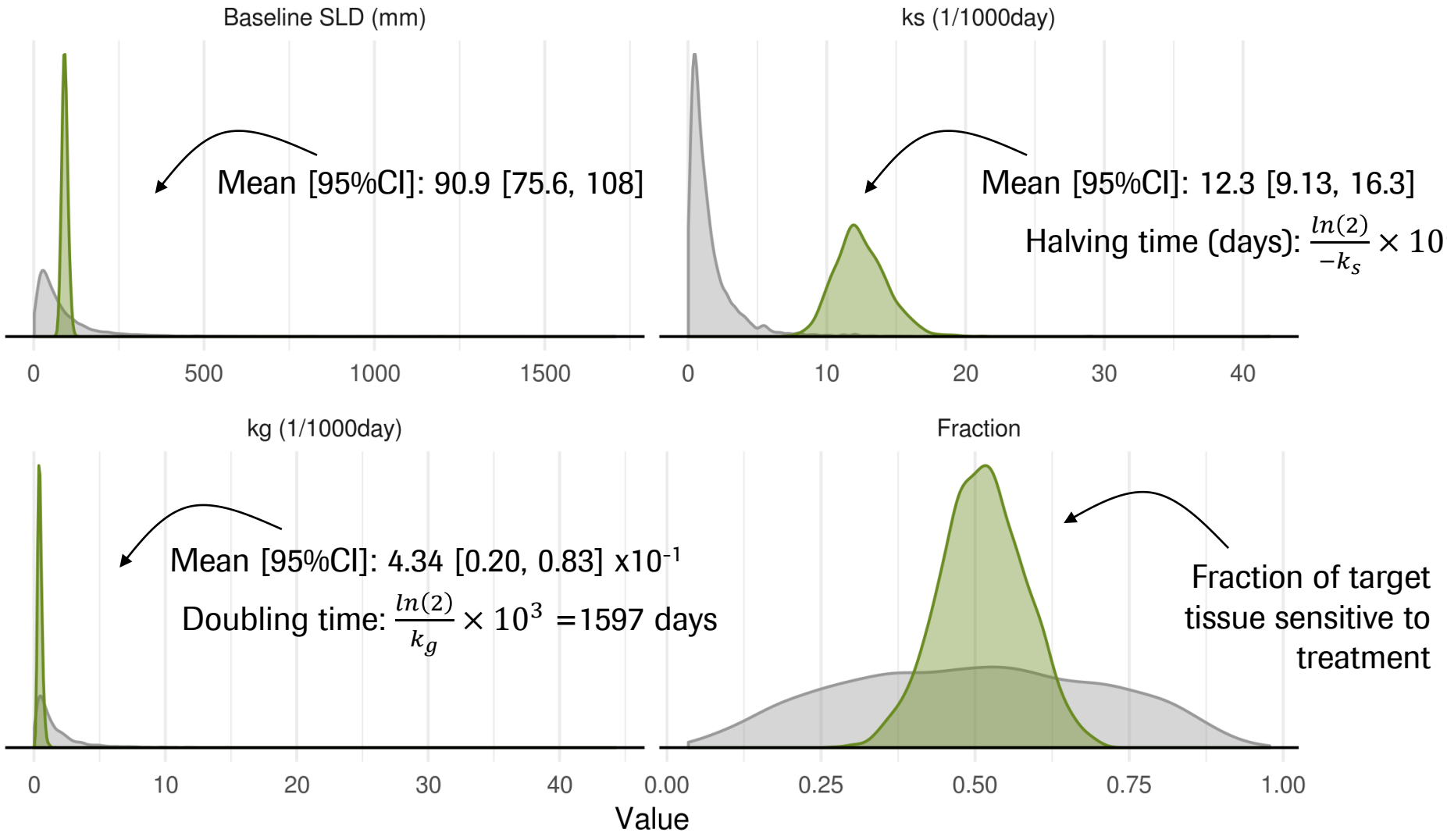
Posterior vs. prior overlap



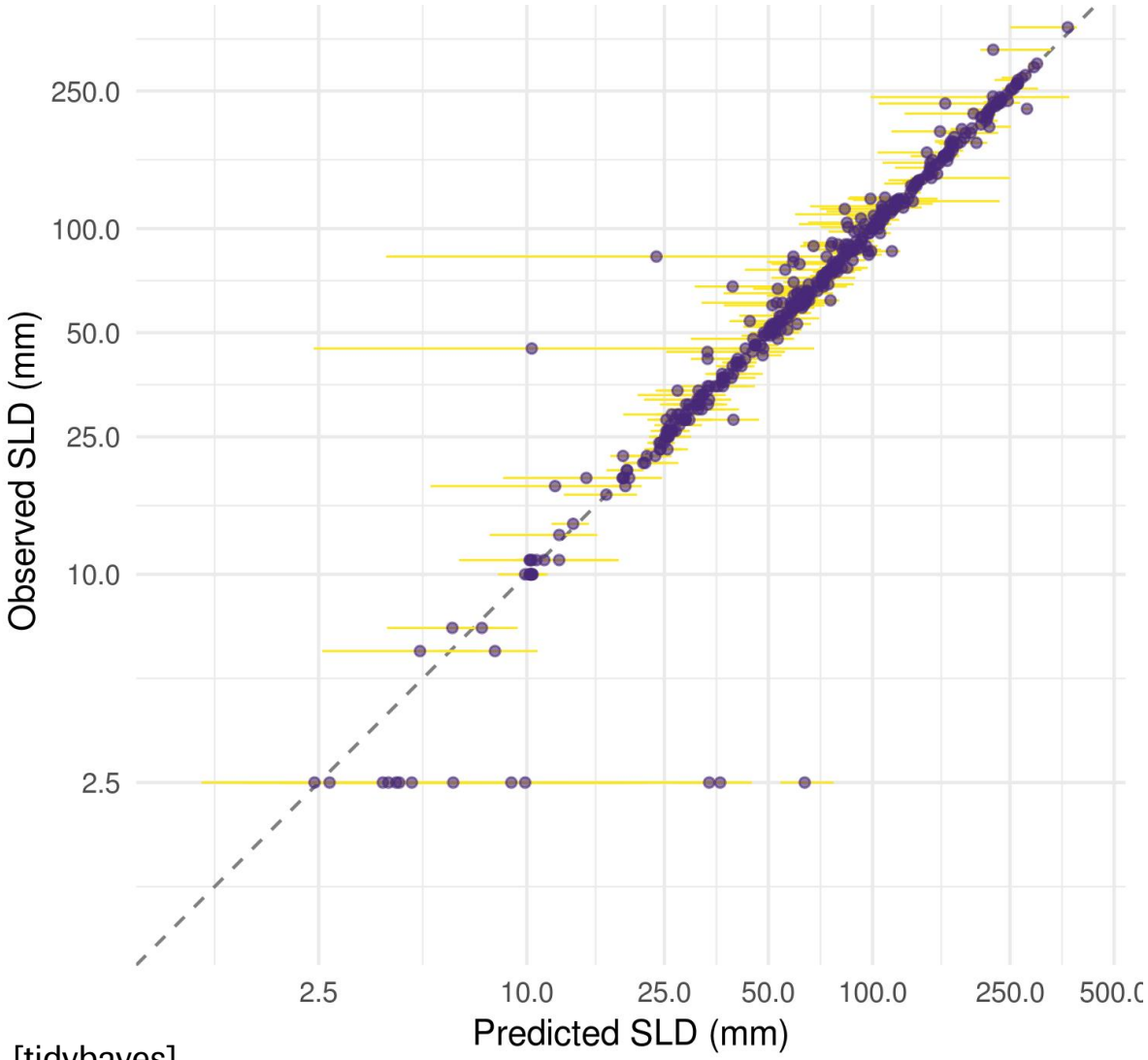
What do we do about this?
Should we adjust our prior?

No … but now, the posterior can become the prior for the analysis of **new data**

Posterior interpretation



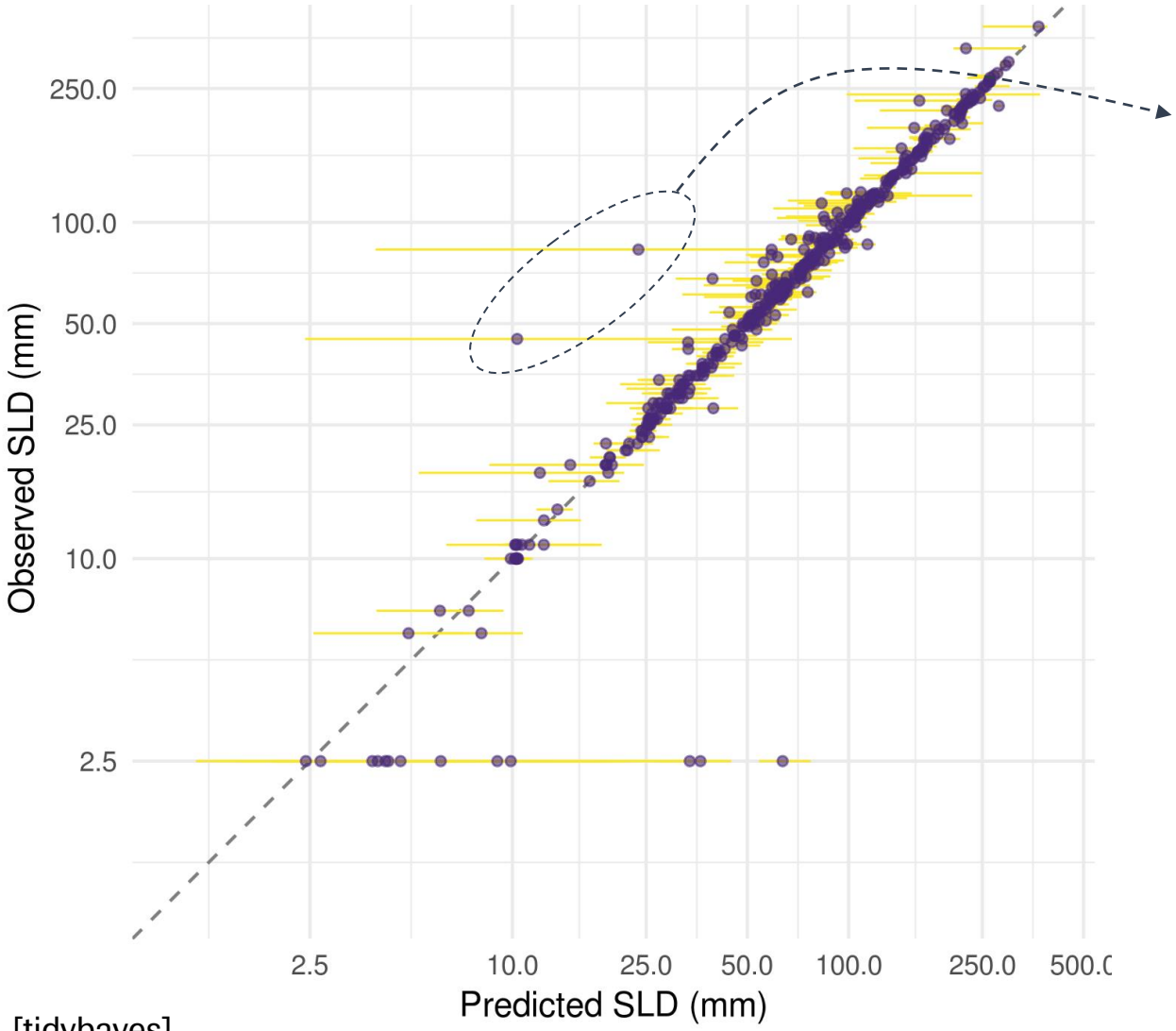
Diagnostic plot: observation vs. prediction

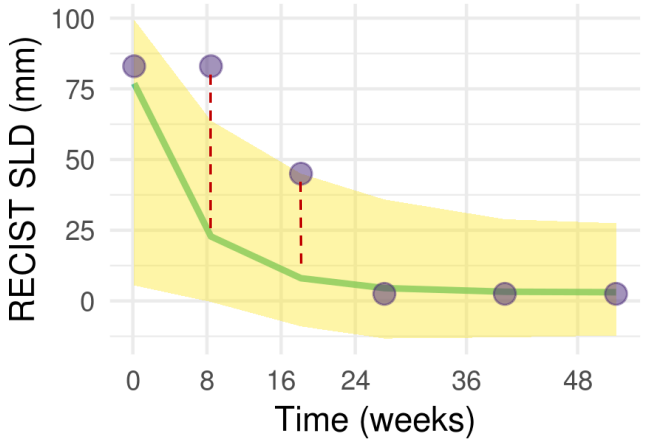


[tidybayes]

Note: Yellow horizontal bars display 95% credible interval around the predicted mean value of the response distribution (*i.e.* excluding residual error).

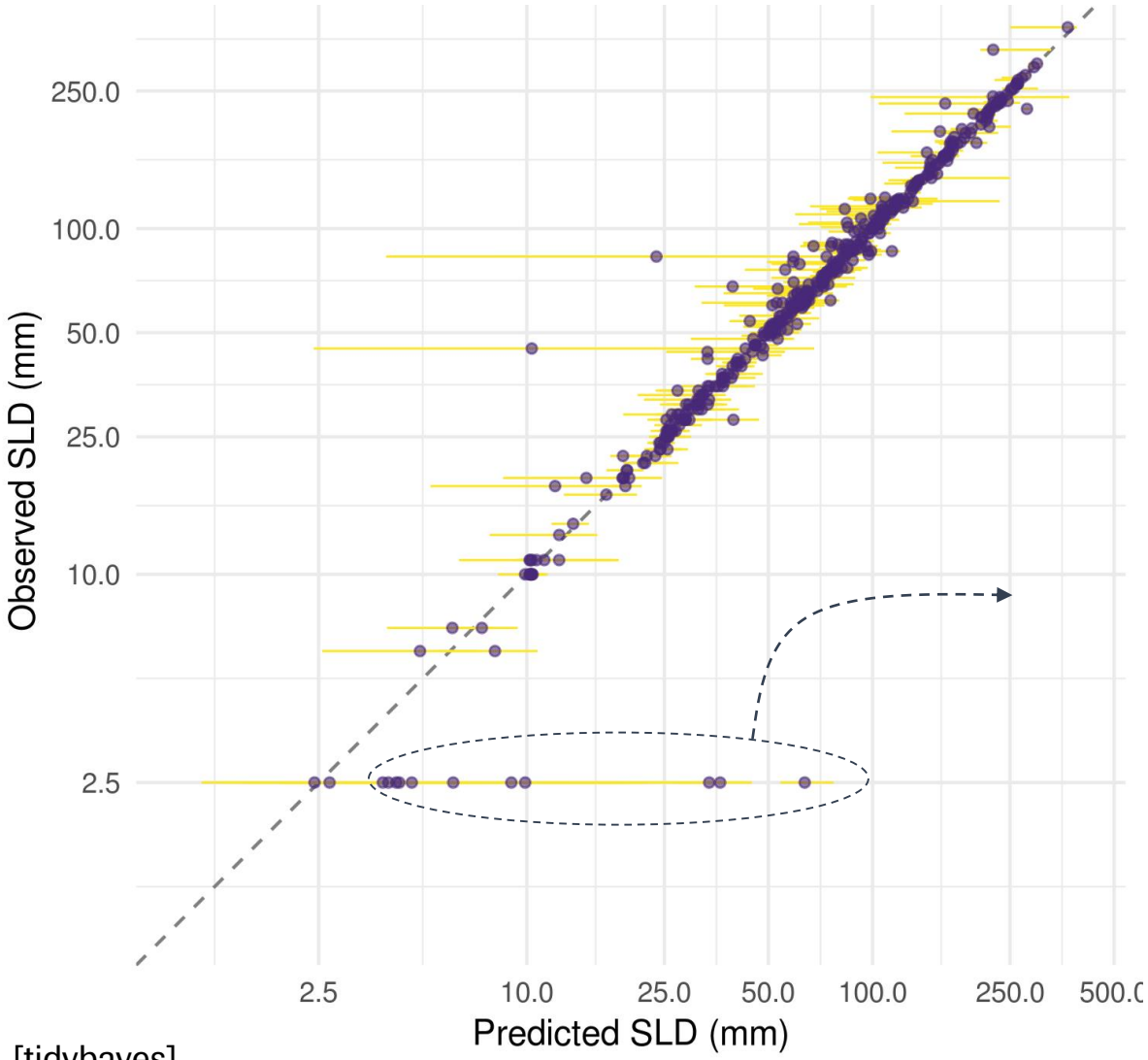
Diagnostic plot: observation vs. prediction





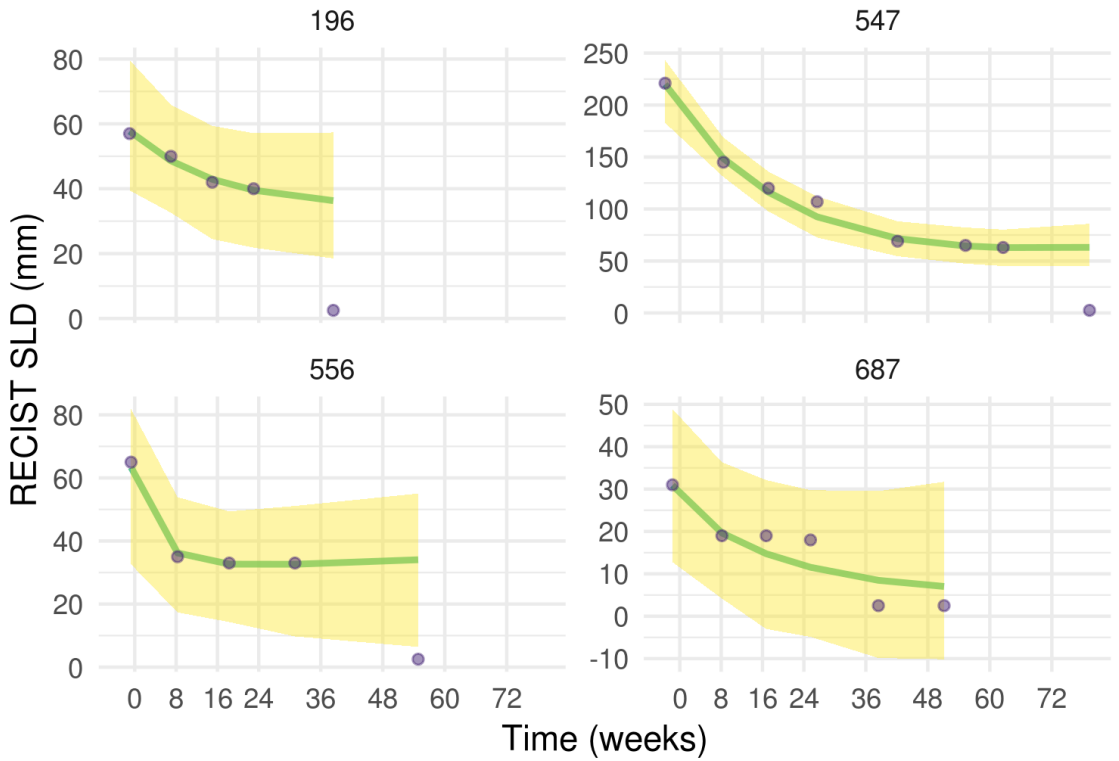
[tidybayes]

Diagnostic plot: observation vs. prediction



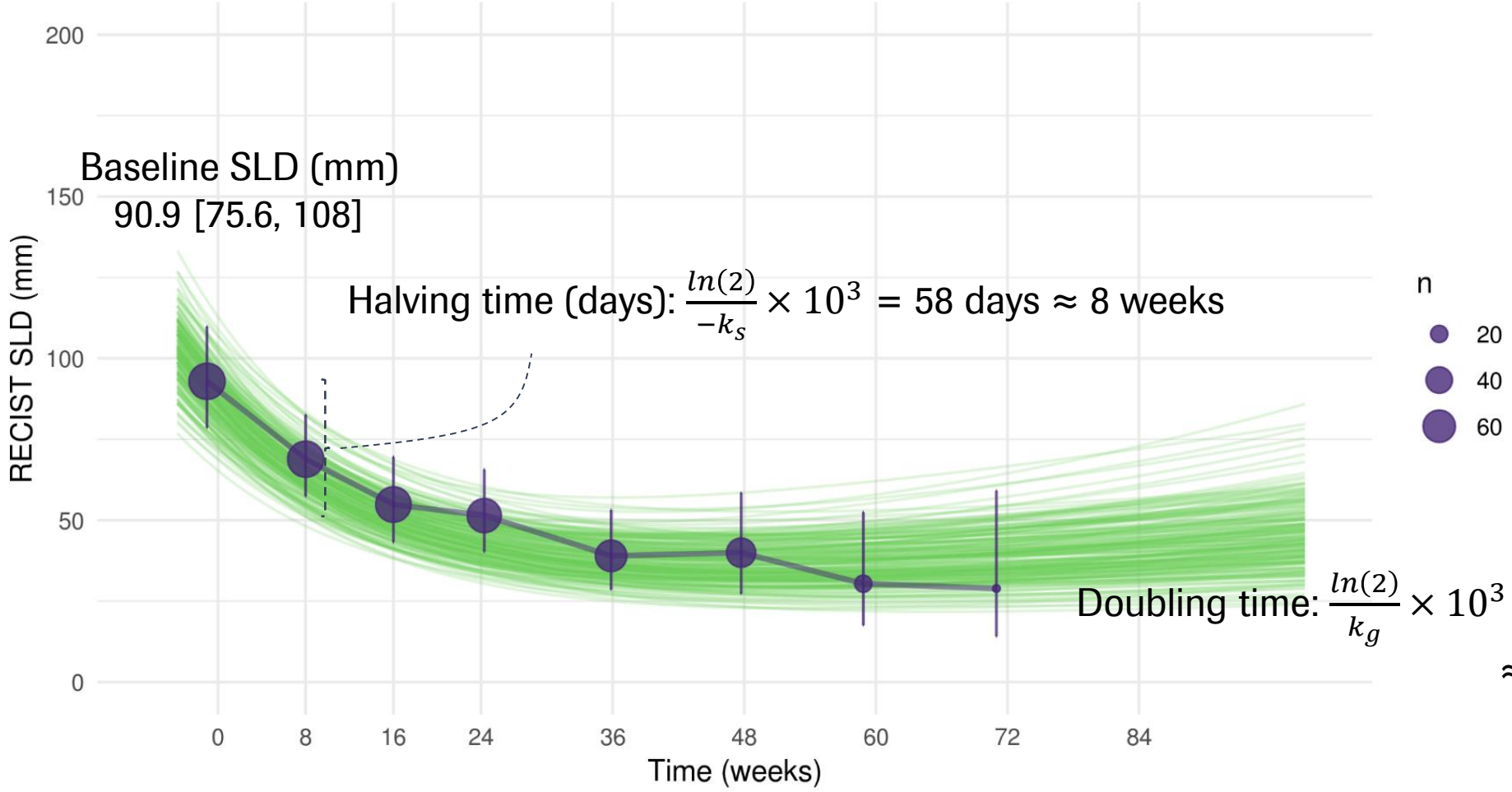
[tidybayes]

The model needs refinement to handle LLoQ



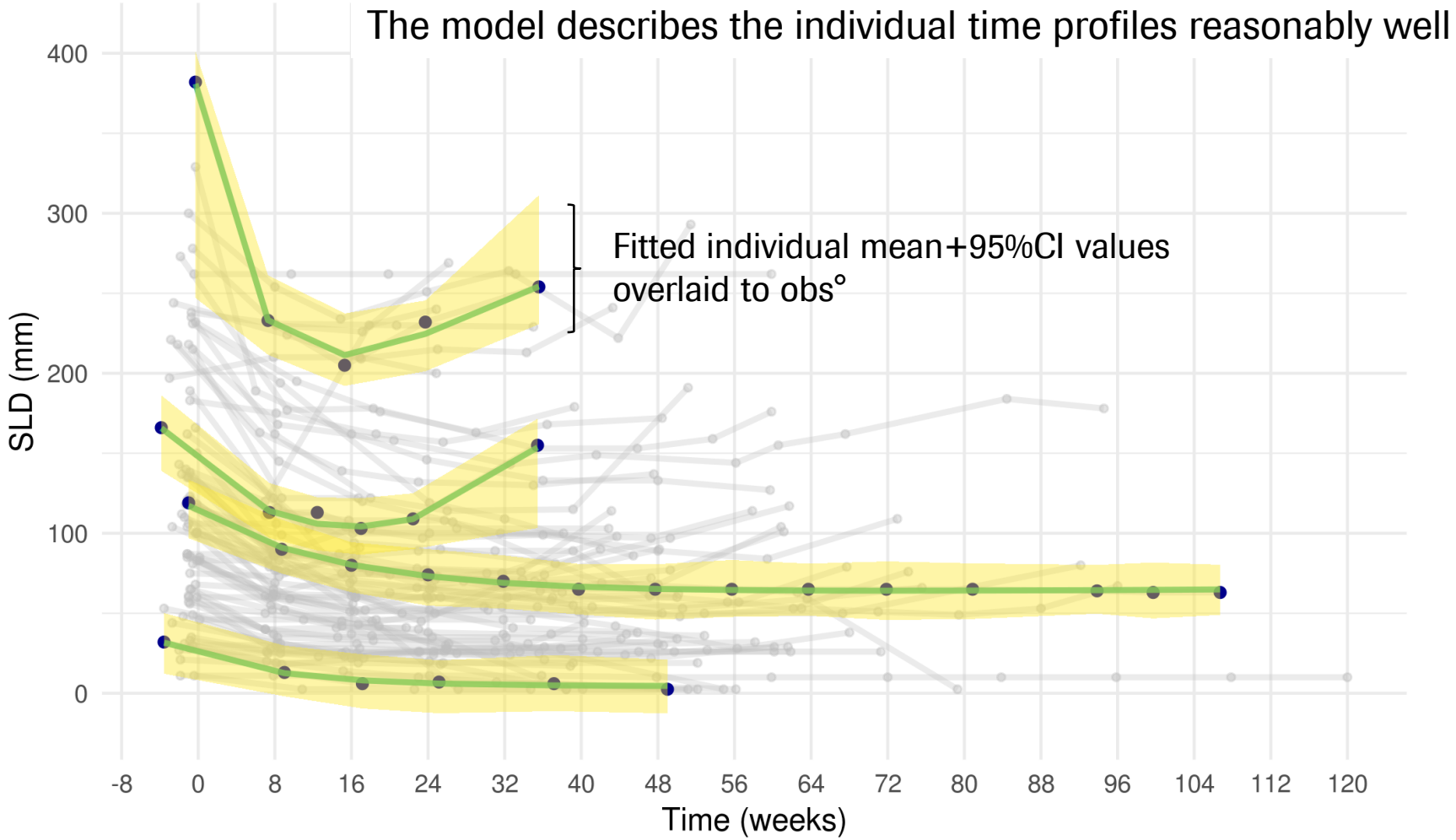
(In-sample) posterior predictive checks

Central tendency



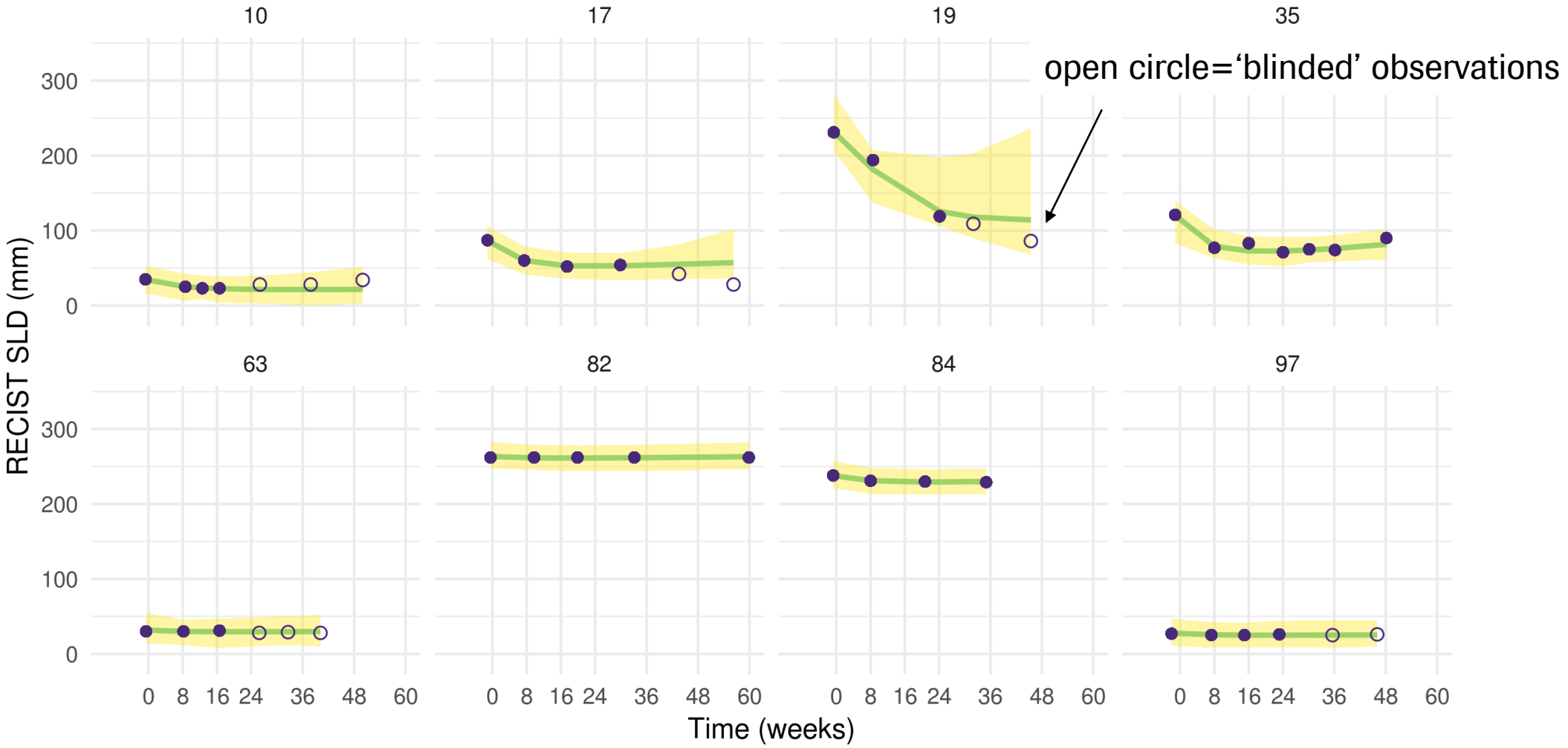
(In-sample) posterior predictive checks

IIV



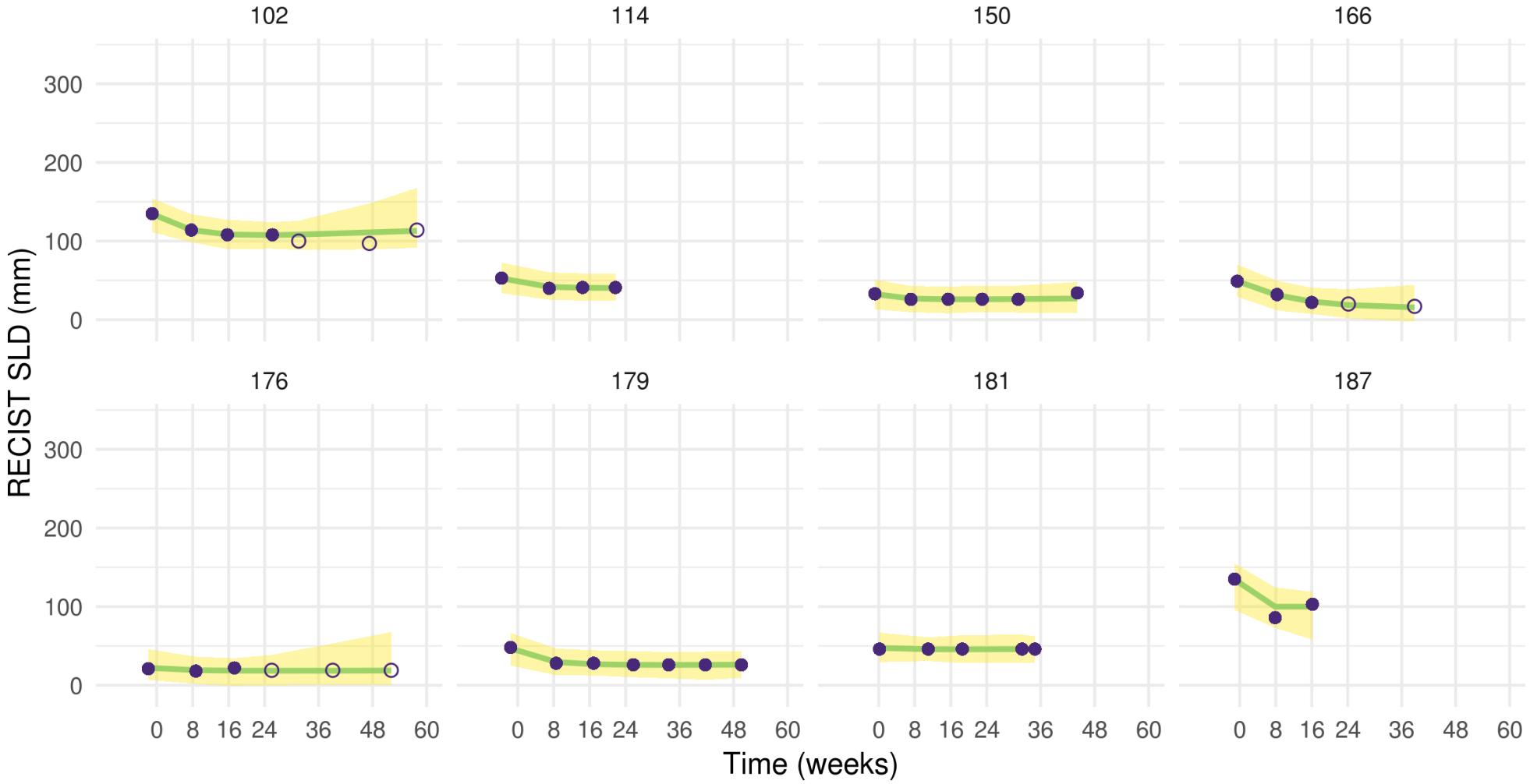
Out of sample predictive check (1/4)

Fit the model with 20% records removed; Confront predictions vs. blinded observations



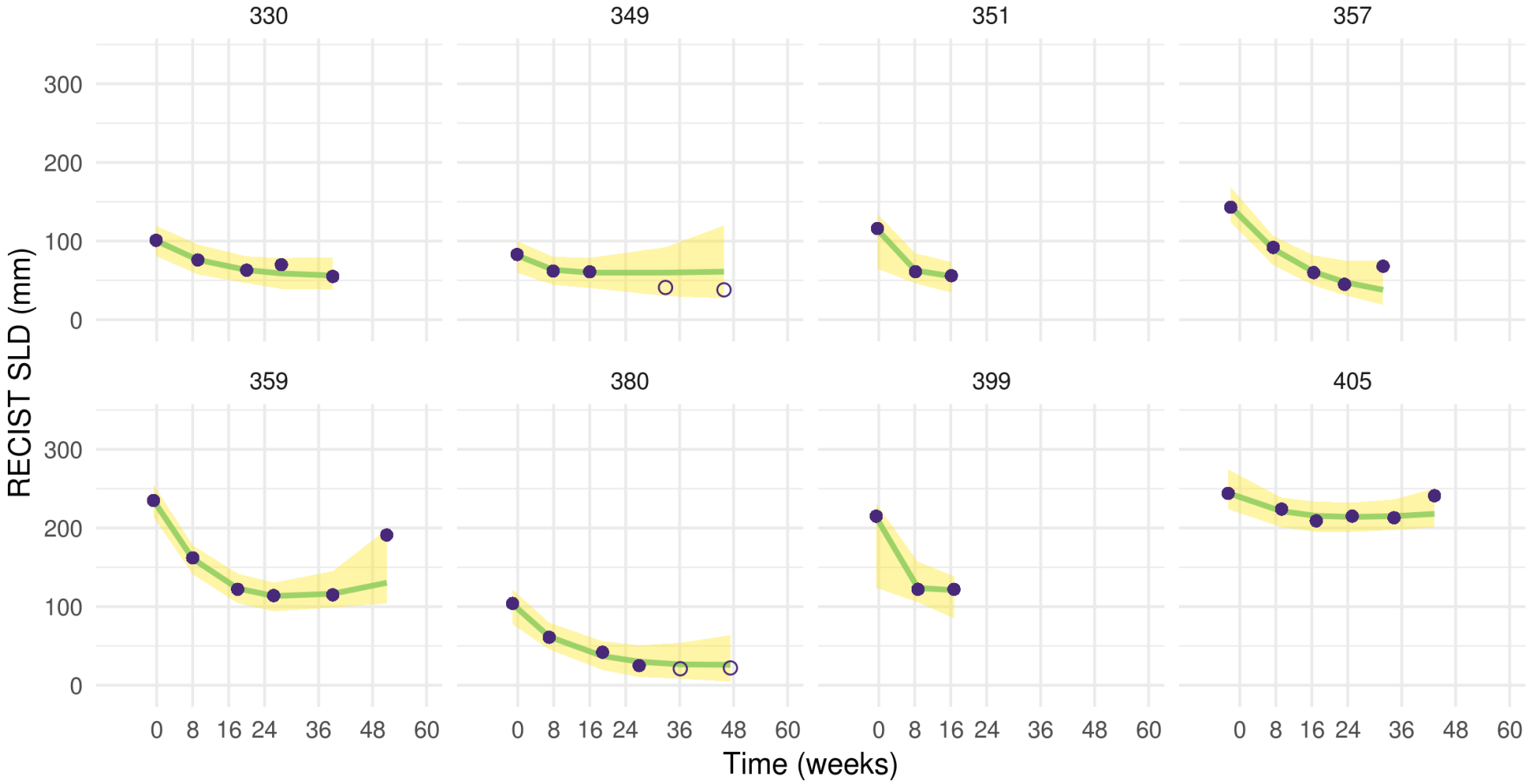
Out of sample predictive check (2/4)

Fit the model with 20% records removed; Confront predictions vs. blinded observations



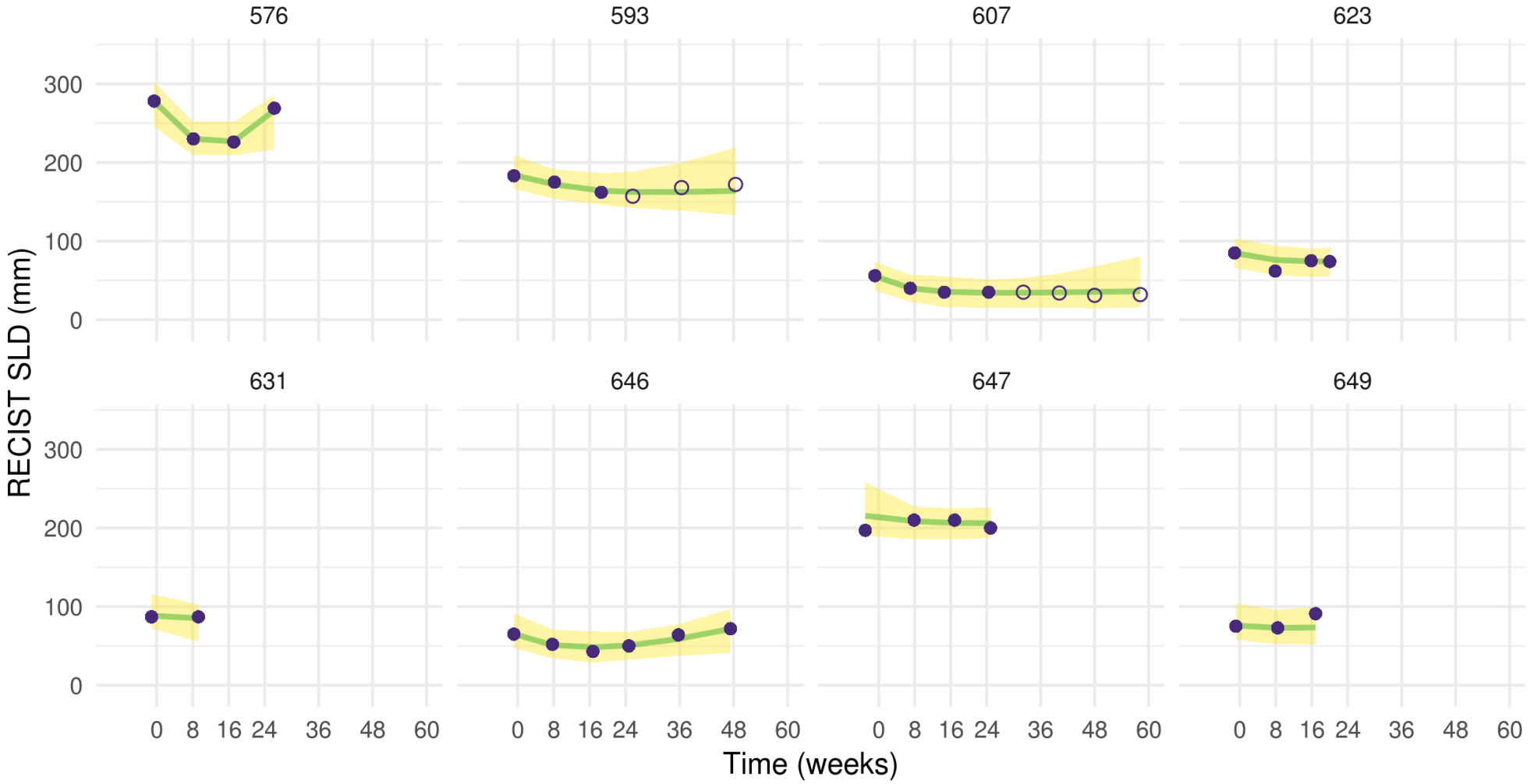
Out of sample predictive check (3/4)

Fit the model with 20% records removed; Confront predictions vs. blinded observations



Out of sample predictive check (4/4)

Fit the model with 20% records removed; Confront predictions vs. blinded observations

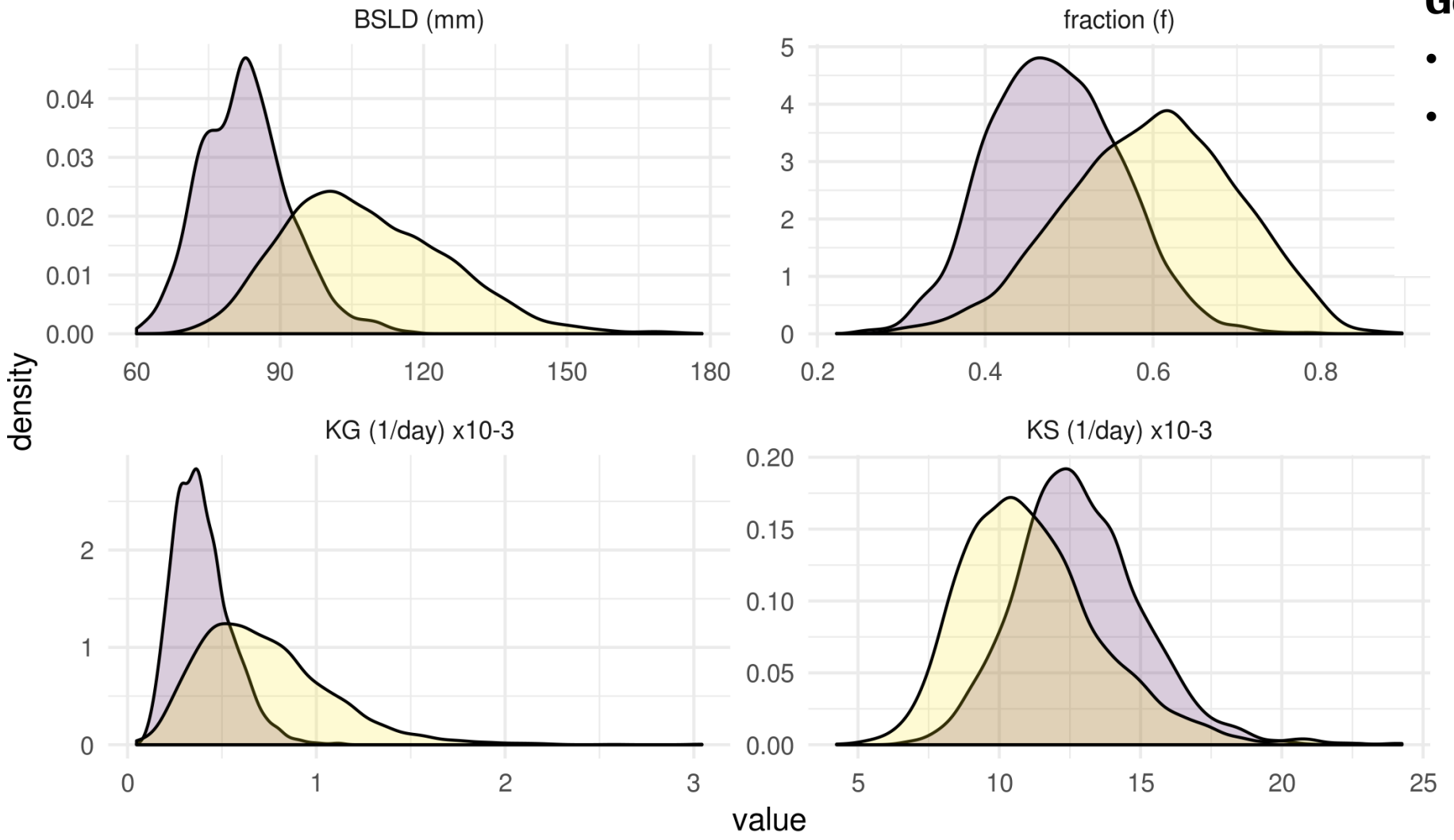


Our goals ➧ Our plan

1. **Summarize** the time trend at the **population** and at the **individual** levels
    - Given a set of **candidate models**, retain the one offering the best **predictive performance**

2. Investigate if/which intrinsic factors (a.k.a. covariates) may contribute to **explaining the IIV**
    - From the best model (among candidates), extract **individual parameter estimates summarizing** the time dynamics of SLD and assess **correlation** with relevant **covariates**

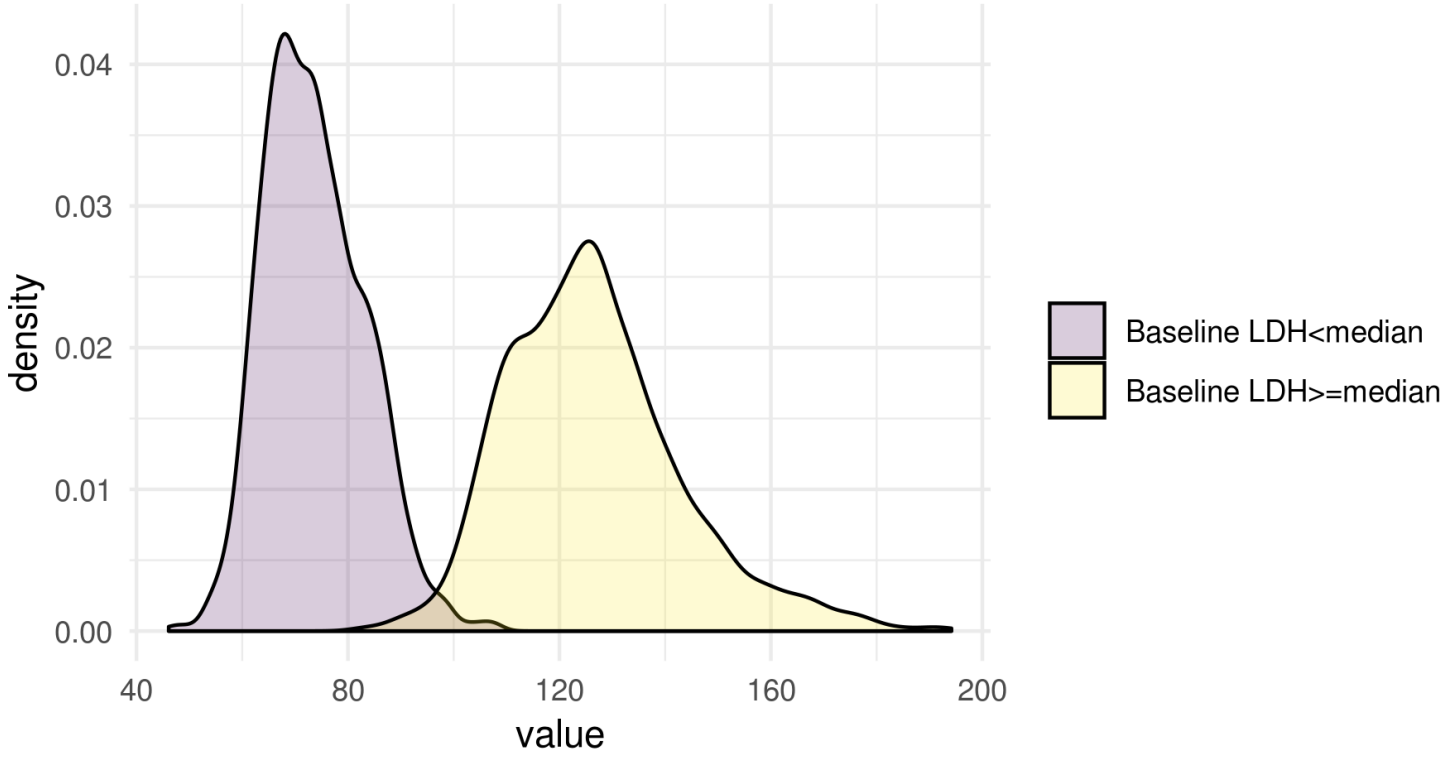
Covariate example 1: Male *vs.* female



GoF

- LOO-IC=3411
- 10-fold CV IC=5054 (39)
  (vs. 4841 w/o Sex)

Covariate example 2: baseline LDH correlated with baseline SLD



GoF

- LOO-IC=3345
- 10-fold CV IC=4634 (37)
  (vs. 4841 w/o LDH)

The packages bundle I used …

Package	Link	Use
brms	https://paul-buerkner.github.io/brms/	Define and fit model, post-processing (summary, plots, predict, …)
loo	https://mc-stan.org/loo/	Compare models (WAIC, LOO IC, k-fold CV IC)
tidybayes	http://mjskay.github.io/tidybayes/	Post-processing
shinystan	https://mc-stan.org/shinystan/	Check model convergence
stanTuner	https://github.com/jhelvy/stanTuner	Find the parameters of prior distributions (normal, beta, inv.gamma)

Conclusion

- Candidate models and priors

  Adequate **transformation** of the parameters is **critical**; Be caution, as the prior and posterior distributions are scaled accordingly

- Recent **wonderful R packages (on top of rstan)** for easy Bayesian models implementation

- Use model for **inference**: ks, kg as efficacy metrics, influence of covariates

Doing now what patients need next