New modifications of Firth's penalized logistic regression

Angelika Geroldinger, Rainer Puhr, Mariana Nold, Lara Lusa, Georg Heinze

24.11.2016, Seminar Graz

This work is supported by FWF under project number I 2276 (“PREMA”).

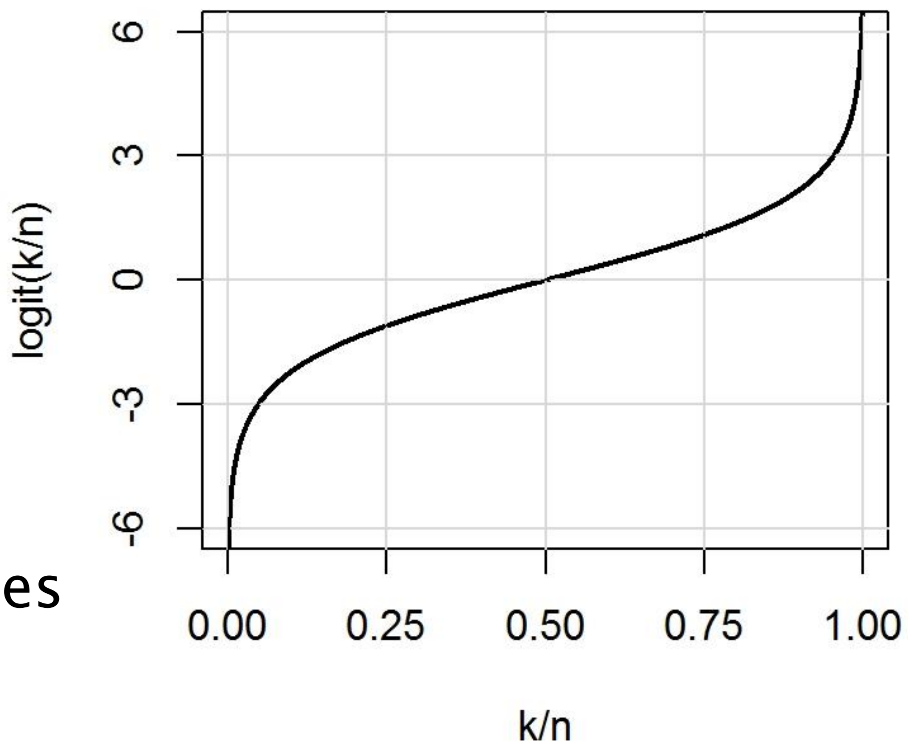
Bias in logistic regression (coefficients)

Consider a model containing only intercept, no regressors:

$$\text{logit}\,(P(Y=1)) = \beta_0.$$

With n observations, k events, the ML estimator of β_0 is given by:

$$\widehat{\beta_0} = \text{logit}\,(\text{k/n}).$$



Since k/n is unbiased,

$\widehat{\beta_0}$ is biased!

Average predicted probabilities are unbiased.

Firth’s penalization

In exponential family models with canonical parametrization **Firth’s penalized likelihood** is given by

$$L^*(\beta) = L(\beta)\,\det(I(\beta))^{1/2},$$

where $I(\beta)$ is the Fisher information matrix and $L(\beta)$ is the likelihood.

Firth’s penalization

- **removes the first-order bias** of ML-estimates of β,
- is **bias-preventive** rather than corrective,
- is available in **Software** packages such as SAS, R, Stata…

Firth's logistic regression (FL)

In logistic regression, the penalized likelihood is given by

$$L^*(\beta) = L(\beta)\,\det(X^t W X)^{1/2}, \text{ with}$$

$$W = \mathrm{diag}(\mathrm{expit}(\mathrm{X_i}\beta)(1 - \mathrm{expit}(\mathrm{X_i}\beta))) = \mathrm{diag}(\pi_i(1 - \pi_i))\ .$$

- penalized estimates always exist.
- W is maximised at $\pi_i = \frac{1}{2}$, i.e. at $\beta = 0$, thus
  - predictions are usually pulled towards $\frac{1}{2}$,
  - coefficients towards zero.

Firth's logistic regression (FL)

For logistic regression with one binary regressor,
Firth's bias correction amounts to adding $1/2$ to each cell:

original

	A	B
0	44	4
1	1	1

Firth's penalization →

augmented

	A	B
0	44.5	4.5
1	1.5	1.5

event rate $= \frac{2}{50} = 0.04$

$OR_{BvsA} = 11$

av. pred. prob. $= 0.054$

$OR_{BvsA} = 9.89$

FLAC

Split the augmented data into the original and pseudo data:

augmented

	A	B
0	44.5	4.5
1	1.5	1.5

→

original

	A	B
0	44	4
1	1	1

+

pseudo

	A	B
0	0.5	0.5
1	0.5	0.5

Define **F**irth's **L**ogistic regression with **A**dditional **C**ovariate as the stratified analysis of the original and pseudo data:

$OR_{BvsA} = 6.63$
av. pred. prob. $= 0.04 =$ observed proportion of events!

FLAC

In the general case (idea):

One can show, that Firth’s penalization is equivalent to ML estimation of augmented data.

FLAC estimates can be obtained by the following steps:

1) Define an indicator variable discriminating between original and pseudo data.
2) Apply ML on the augmented data including the indicator.

✓ unbiased pred. probabilities

FLIC

Firth's **L**ogistic regression with **I**ntercept **C**orrection:

Modify the intercept in penalized estimates such that the average pred. probabilities becomes equal to the observed proportion of events.

✓ unbiased pred. probabilities

effect estimates are the same as in Firth's logistic regression

Simulation study

We want to investigate

- individual predicted probabilities by FLIC and FLAC
- effect estimates by FLAC
- compare not only against ML and FL but also against
    - weakened Firth's penalization, with $L(\beta)^* = L(\beta)\,\det(X^t W X)^\tau$, $\tau < 1/2$, (WF)
    - ridge regression, (RR)
    - penalization by log-F(1,1) priors, (LF)
    - penalization by Cauchy priors with scale parameter=2.5. (CP)

Simulation study: the set-up

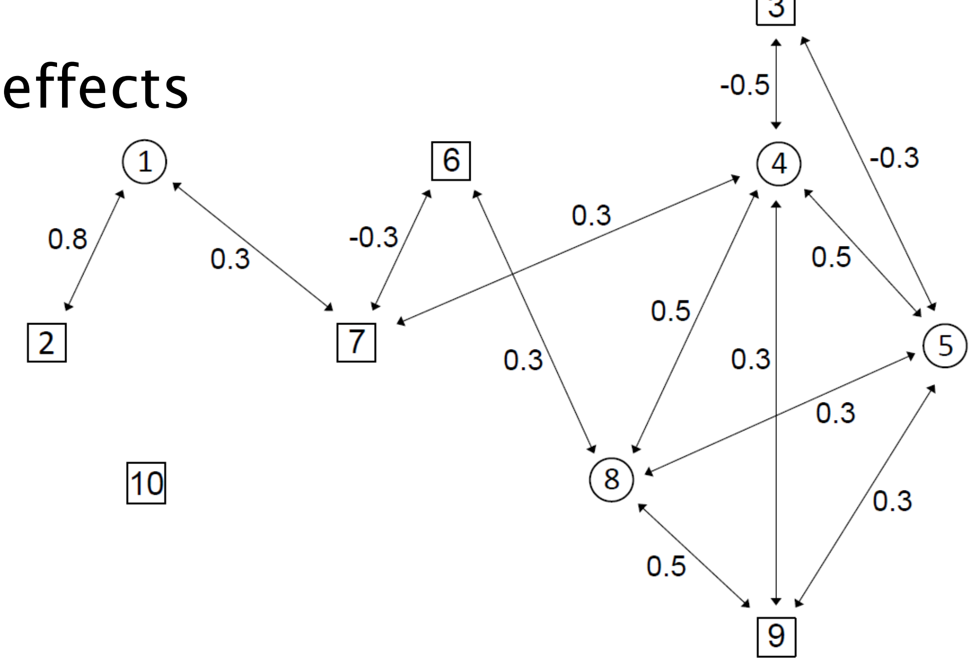
We simulated 1000 data sets for 45 scenarios with:

- 500, 1000 or 1400 observations,
- event rates of 1%, 2%, 5% or 10%
- 10 covariables (6 cat., 4 cont.), see Binder et al., 2011
- none, moderate and strong effects

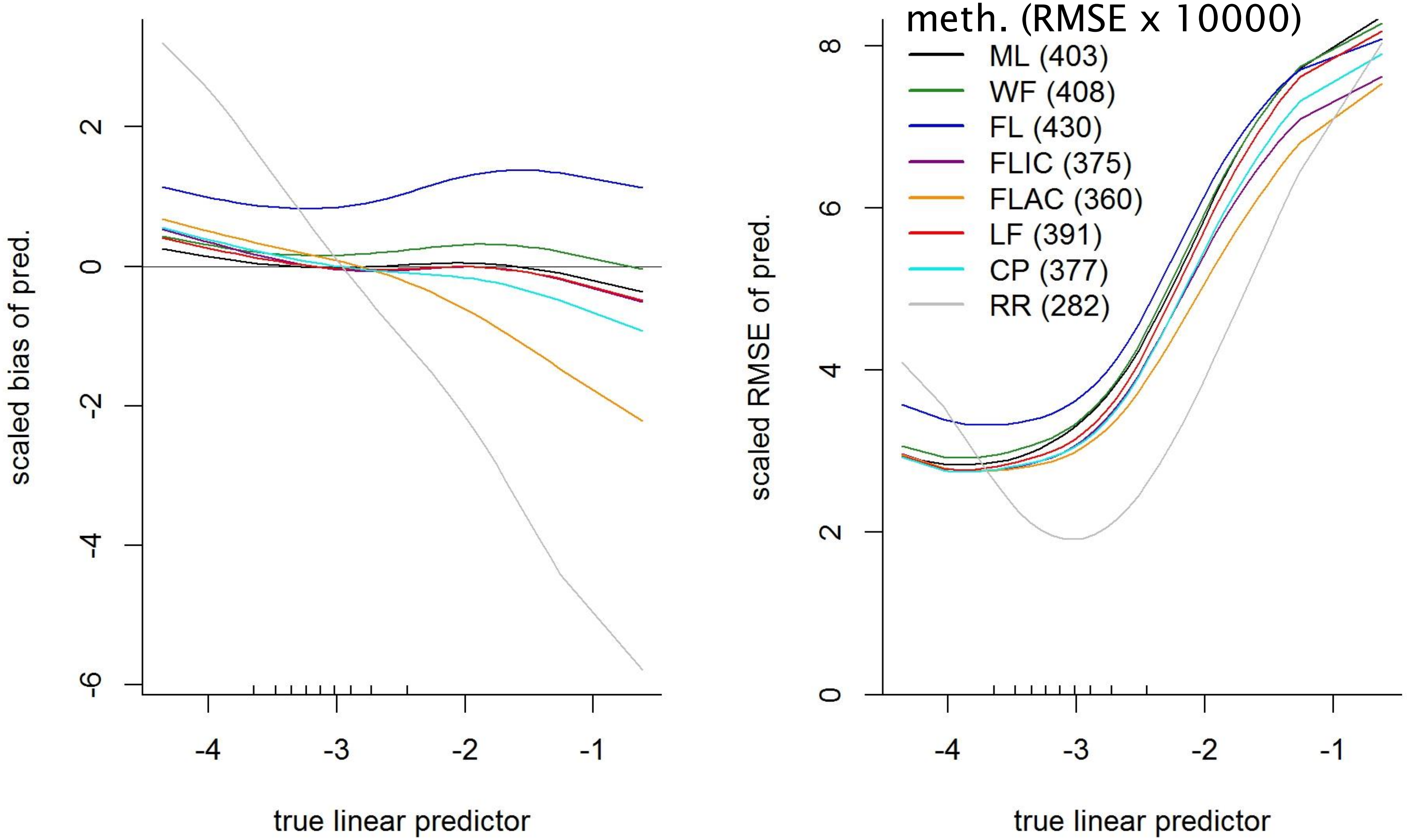
Main evaluation criteria:

Bias and RMSE of

- pred. prob. and
- effect estimates

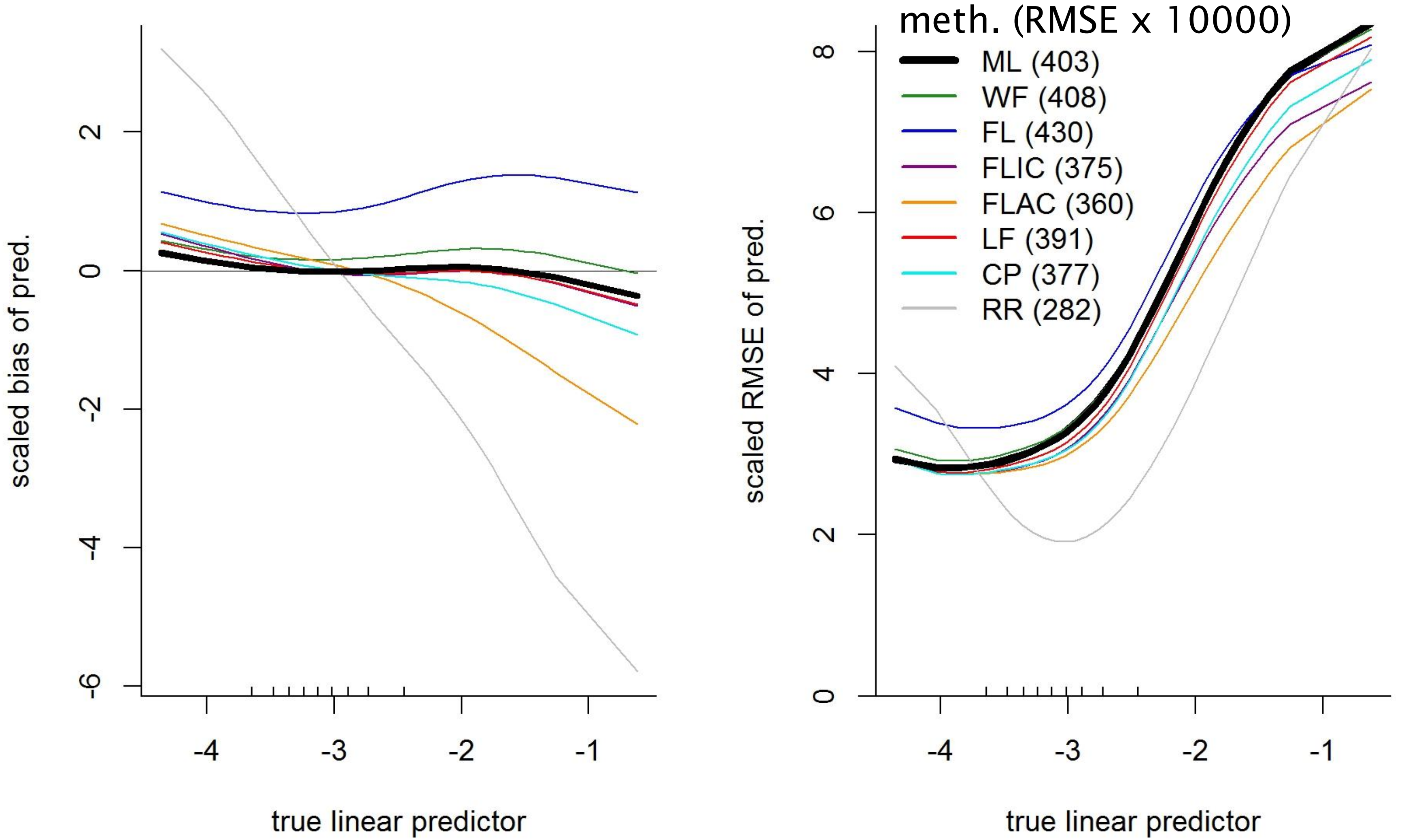


Pred. probabilities by true linear predictor



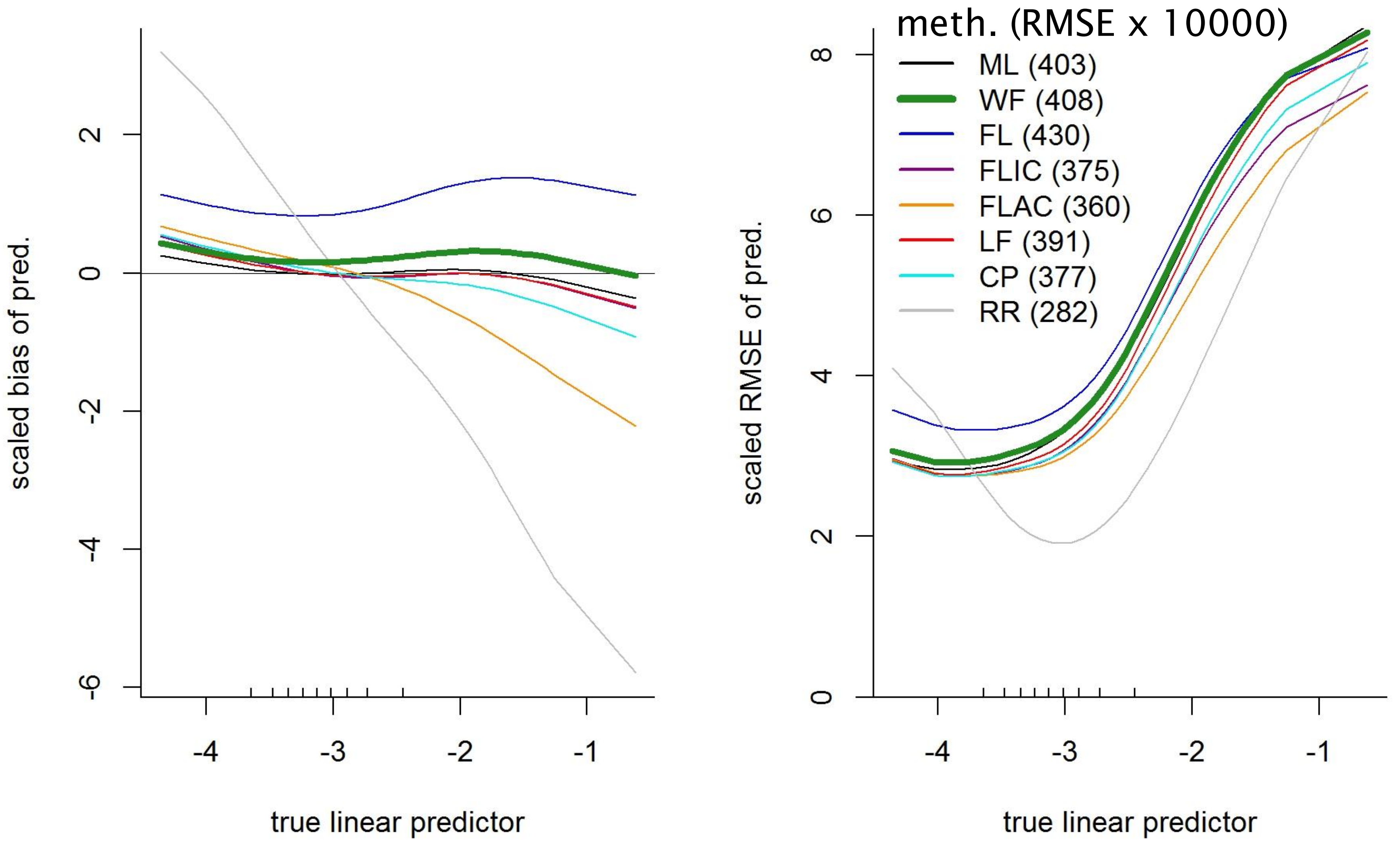
(sample size=500, prop. of events= 5%, small effect size)

Pred. probabilities by true linear predictor



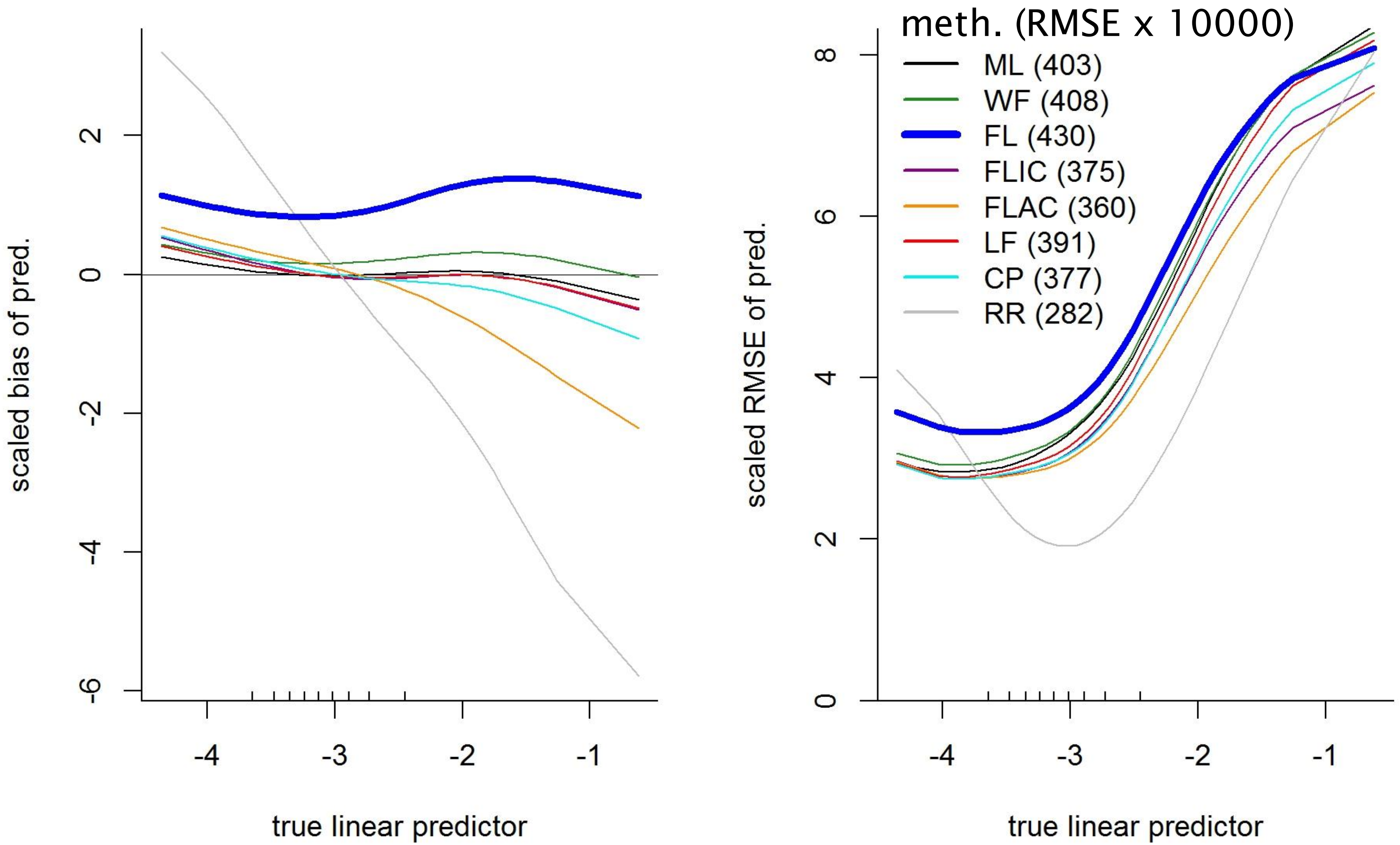
(sample size=500, prop. of events= 5%, small effect size)

Pred. probabilities by true linear predictor



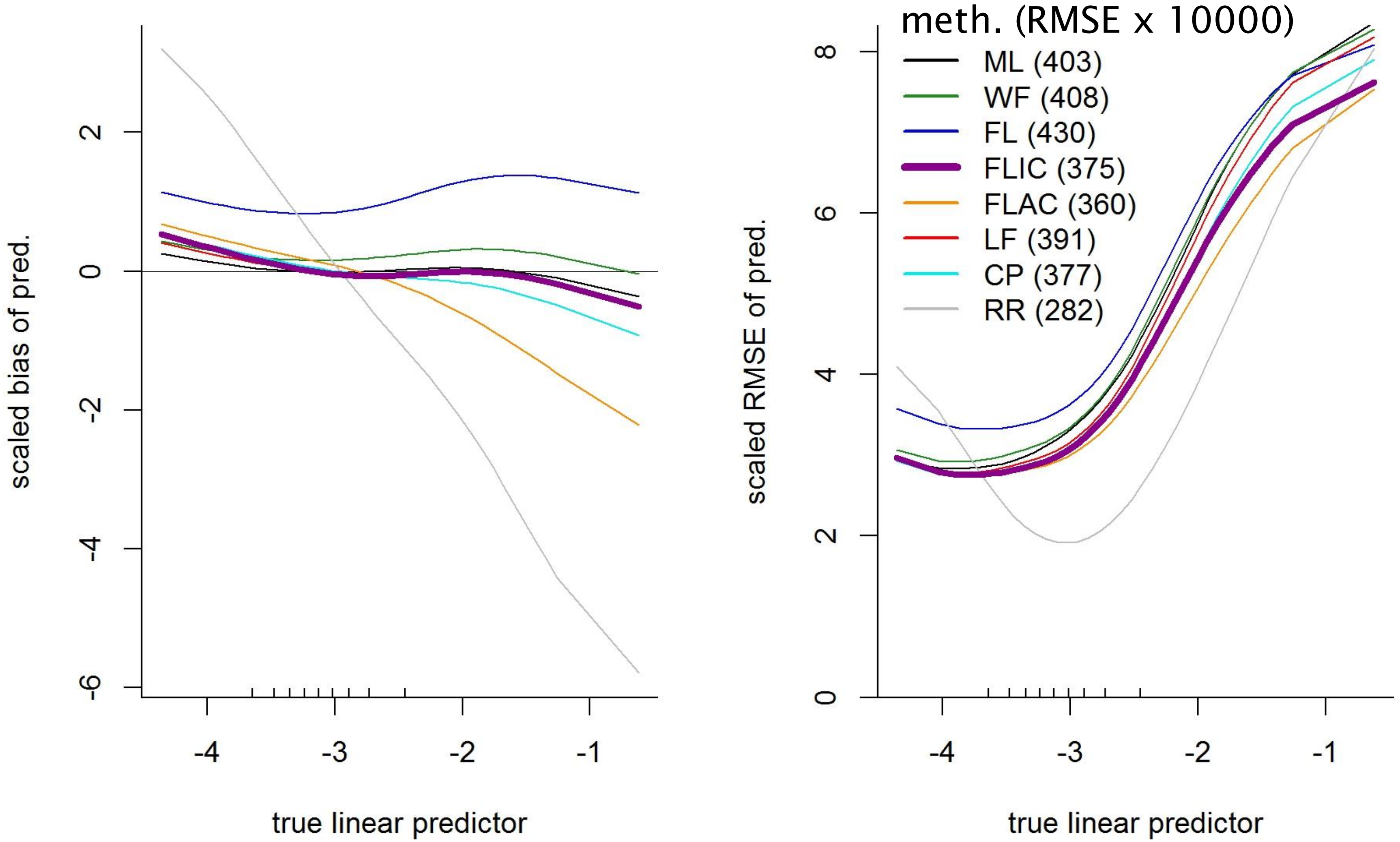
(sample size=500, prop. of events= 5%, small effect size)

Pred. probabilities by true linear predictor



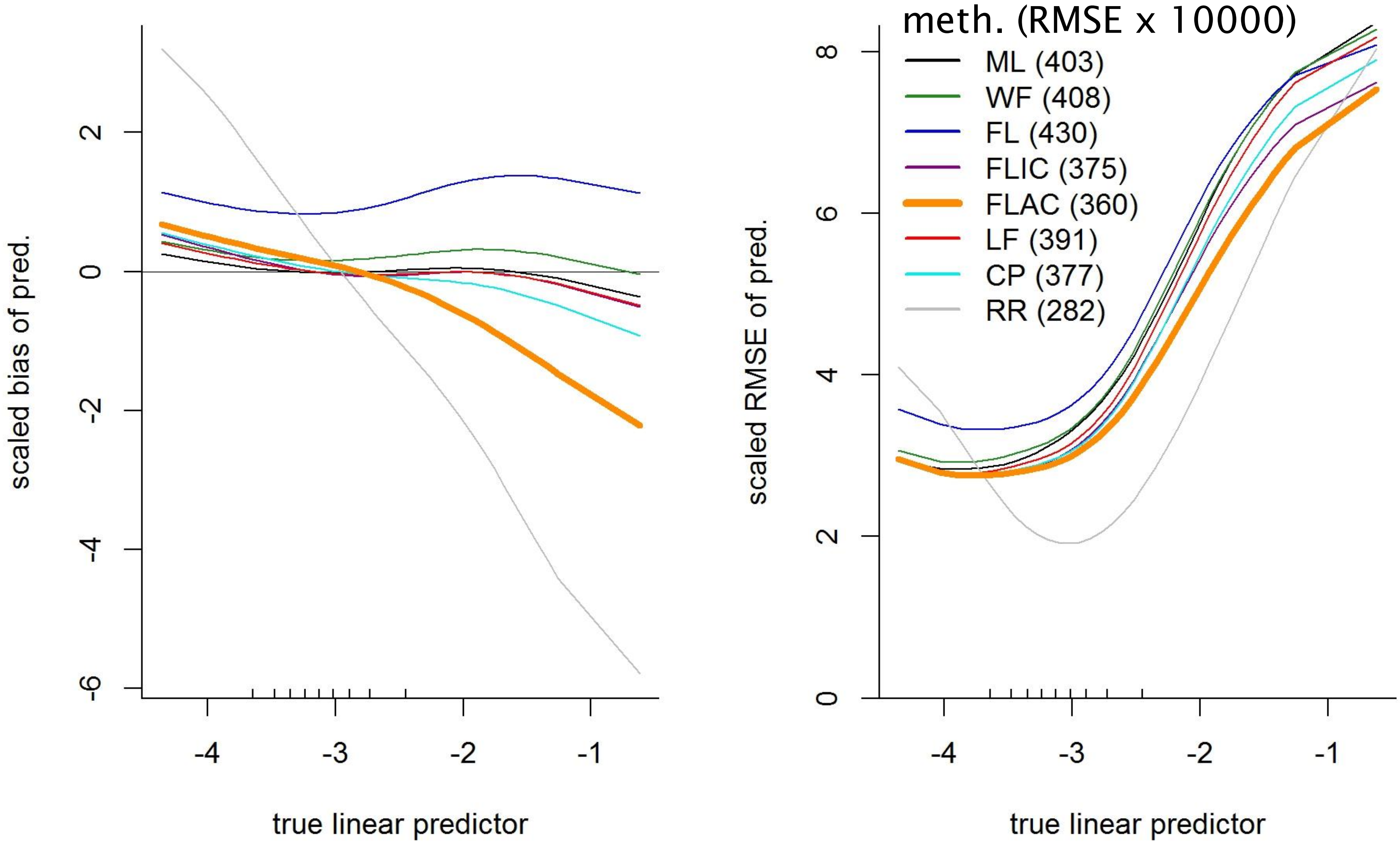
(sample size=500, prop. of events= 5%, small effect size)

Pred. probabilities by true linear predictor



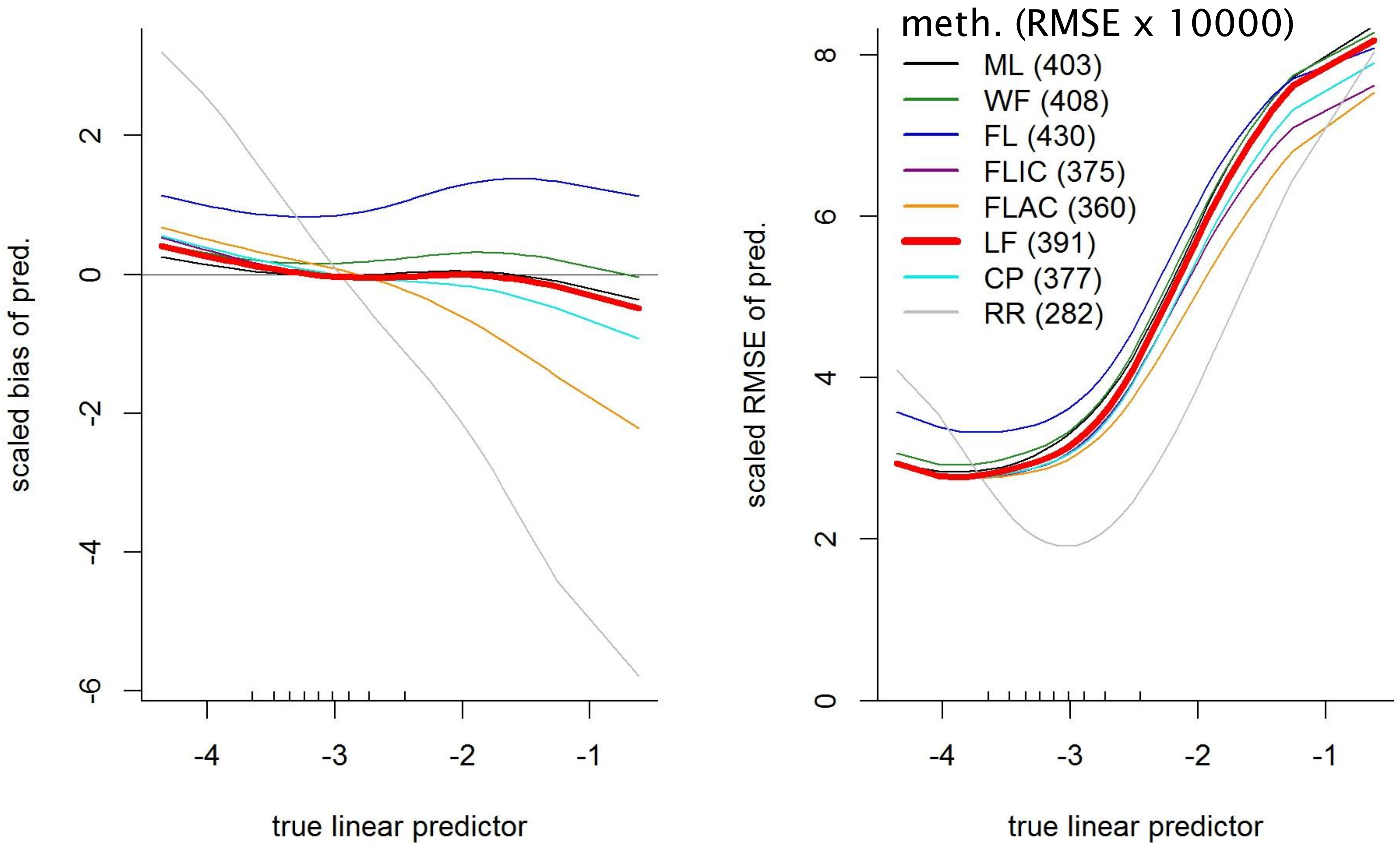
(sample size=500, prop. of events= 5%, small effect size)

Pred. probabilities by true linear predictor



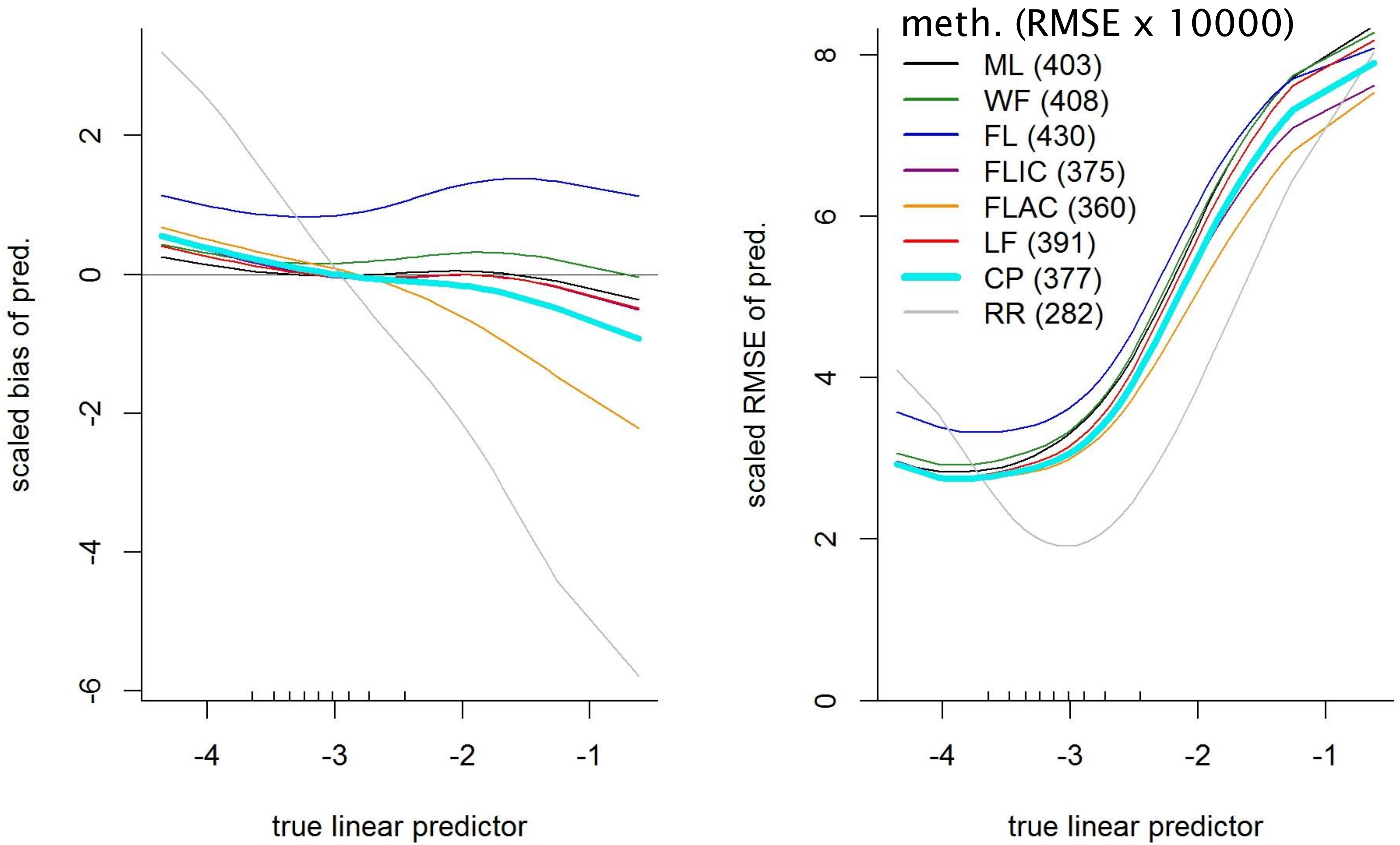
(sample size=500, prop. of events= 5%, small effect size)

Pred. probabilities by true linear predictor



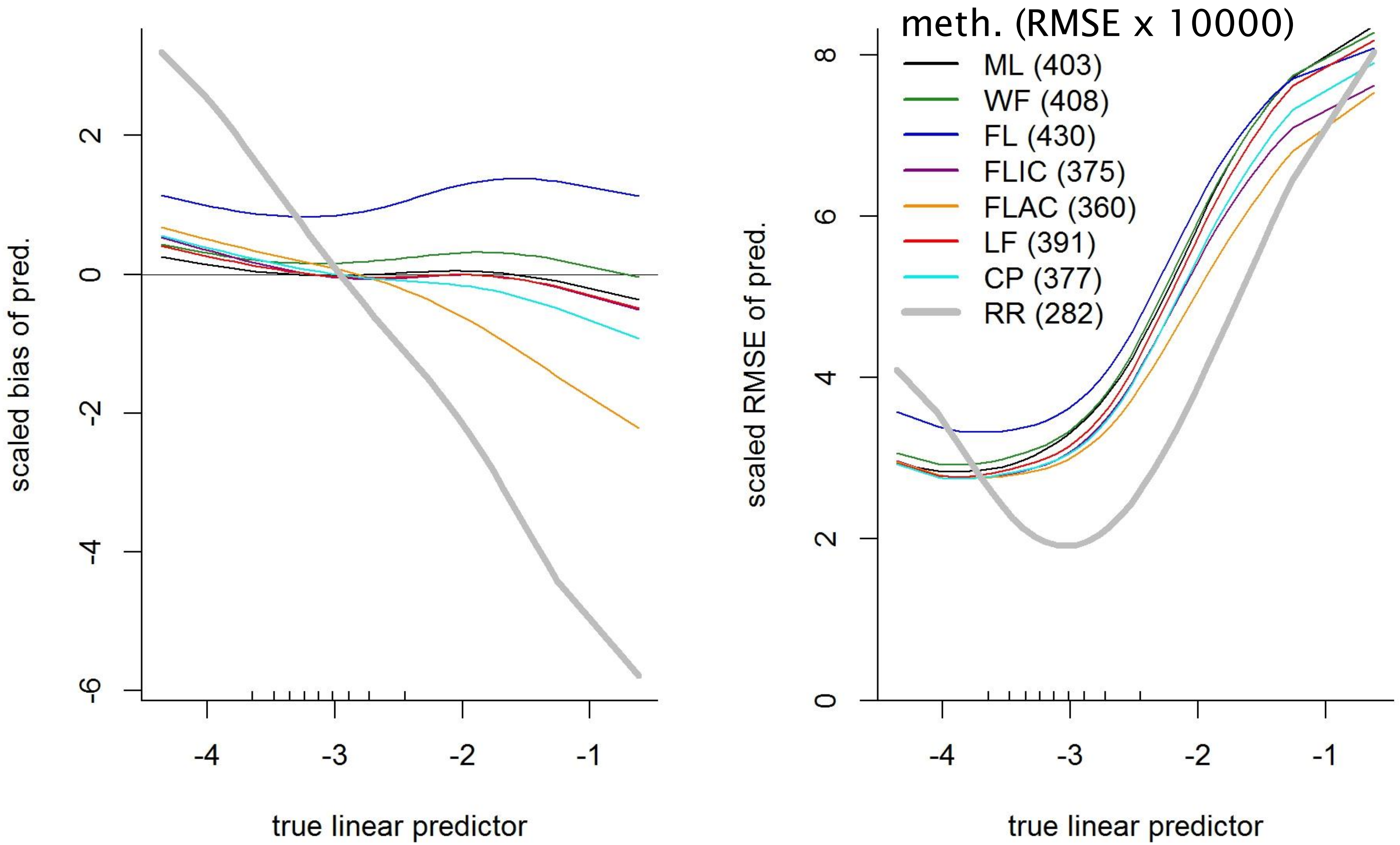
(sample size=500, prop. of events= 5%, small effect size)

Pred. probabilities by true linear predictor



(sample size=500, prop. of events= 5%, small effect size)

Pred. probabilities by true linear predictor



(sample size=500, prop. of events= 5%, small effect size)

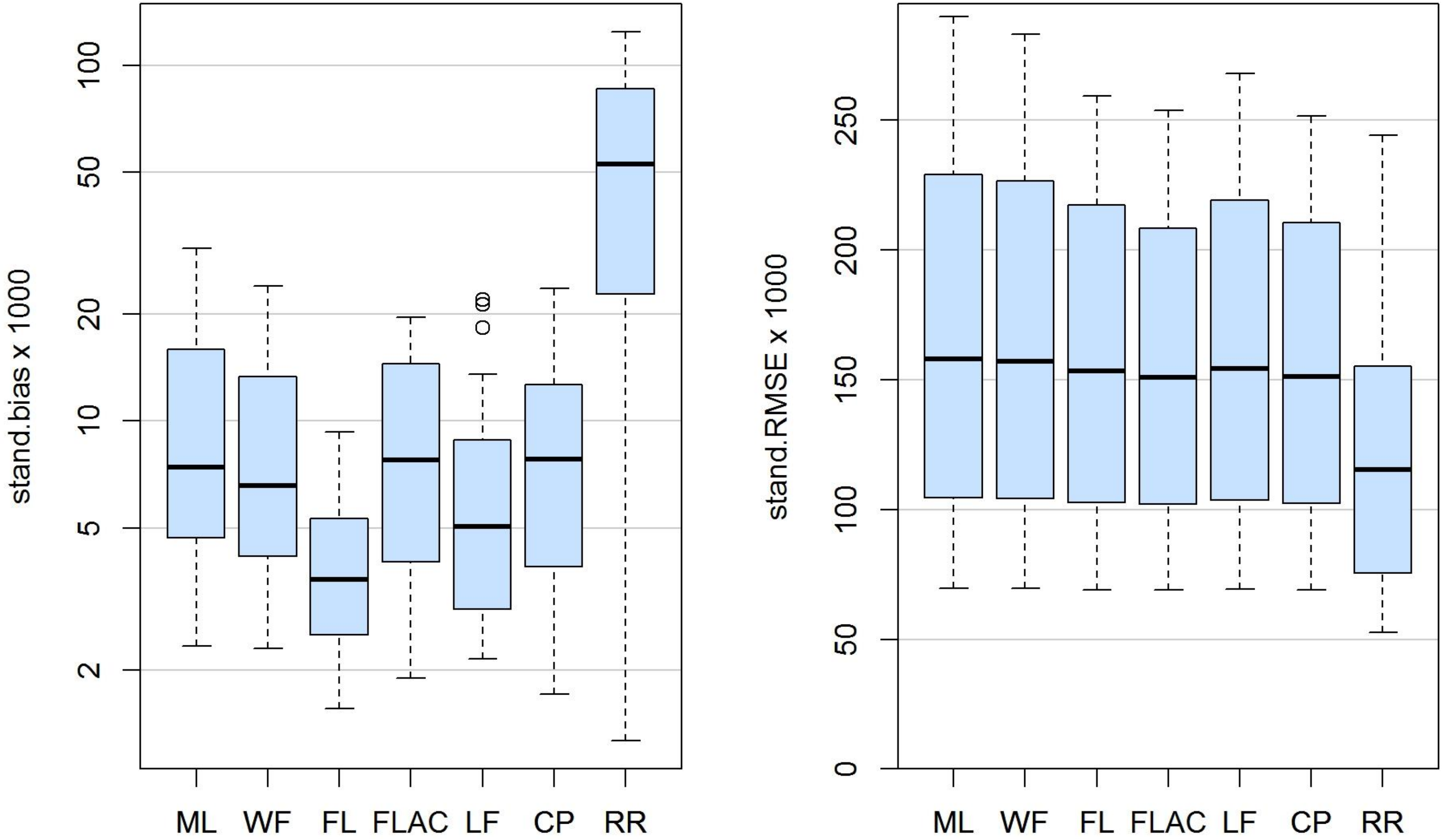
Coefficients

Absolute bias and RMSE of standardized coefficients, averaged over all 10 covariables excluding the intercept:

N	E(y)	method	bias (×1000) effect size			RMSE (×1000) effect size		
			0	0,5	1	0	0,5	1
500	0.05	ML	23	17	29	277	266	288
		WF	19	14	21	272	261	281
		FL/FLIC	7	5	9	253	244	259
		FLAC	17	16	16	239	235	252
		LF	22	10	12	265	252	266
		CP	18	14	24	245	238	251
		RR	3	109	124	78	166	244

Coefficients

Similar patterns can be observed across all 45 scenarios:



Conclusions

- For rare events, FL-predictions are severely biased (relative bias of up to 20% in our simulations).
- Both, FLIC and FLAC improved on predictions by FL, with identical effect estimates or effect estimates of lower RMSE.
- RR outperformed all other methods with respect to RMSE of coefficients and predictions, but introduces bias towards 0. Confidence intervals?

Based on our simulations, if one is interested in effect estimates and predictions, we recommend to use

  - RR (whenever confidence intervals are not needed)
  - FLAC as a compromise between optimization of bias and RMSE.

Literature

- Binder H, Sauerbrei W and Royston P. Multivariable Model-Building with Continuous Covariates: Performance Measures and Simulation Design 2011. Technical Report FDM-Preprint 105, University of Freiburg, Germany.
- Elgmati E, Fiaccone RL, Henderson R and Matthews JNS. Penalised logistic regression and dynamic prediction for discrete-time recurrent event data. Lifetime Data Analysis 2015; 12(4): 542-560.
- Firth D. Bias reduction of maximum likelihood estimates. Biometrika 1993; 80(1): 27-38.
- Gelman A, Jakulin A, Pittau M and Su YS. A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models. *The Annals of Applied Statistics* 2008; 2(4):1360–1383.
- Greenland S and Mansournia M. Penalization, Bias Reduction, and Default Priors in Logistic and Related Categorical and Survival regressions. *Statistics in Medicine* 2015; 34(23):3133-3143.
- Heinze G and Schemper M. A solution to the problem of separation in logistic regression. Statistics in Medicine 2002; 21(16): 2409-2419.
- Kosmidis I. Bias in parametric estimation: reduction and useful side-effects. WIRE Computational Statistics 2014; 6(3): 185-196.
- Puhr R, Heinze G, Nold M, Lusa L and Geroldinger A. Predicting rare events with penalized logistic regression. Work in progress.