BNR Workshop, INI, Cambridge
August 9, 2007

Postulating monotonicity in nonparametric Bayesian regression

Elja Arjas
Department of Mathematics and Statistics,
University of Helsinki
and
National Public Health Institute (KTL)

Based on ongoing joint work with Olli Saarela (KTL)

Outline of talk

- Motivation
- Previous approaches
- Unconstrained model: prior, illustration, MCMC
- Model with monotonicity constraints
- Illustration of the method with real epidemiological data
- Summary, conclusions

Warning

- No mathematical results in this talk!

Motivation

- In areas such as epidemiology, strong parametric assumptions are often imposed on the form of the regression function describing covariate effects.

- This is done as a modelling convention, typically without real support from contextual substantive arguments, evidence coming from earlier studies, or careful diagnostics afterwards.

Motivation (2)

- Can one relax assumptions based on e.g. linearity/additivity or multiplicativity (proportional hazards, proportional odds)?

- Completely nonparametric estimation of regression surface may give a too complex (”noisy”) picture of the true functional relationship as small disturbances in the data affect the function estimate.

Motivation (3)

- However: Often there are reasonable grounds to assume a monotonic relationship between a covariate and the considered response.

- (cf. remarks David Dunson made at the end of his talk on Monday: ”In biostatistical applications, we actually often know a lot ...)

Previous approaches

- For example, Arjas & Gasbarra (1994), Schell & Singh (1997) and Holmes & Heard (2003) have applied a monotonicity assumption coupled with a piecewise constant model with random change points for *curve estimation in one dimension*.

- Can one find computationally feasible generalisations to *multiple dimensions,* without making additional parametric assumptions?

Unconstrained model

- Consider a model for log hazard rate
  $\log(\lambda) = \theta(x_1, ..., x_L) + \beta_1 z_1 + ... + \beta_M z_M$,
  where x are factors to be handled non-parametrically (can be time scales or covariates) and z are factors for which linearity and additivity is assumed.

- Function $\theta(x_1, ..., x_L)$ is handled by applying a set of random change points for the covariate axes and assuming constancy of θ within the resulting multivariate intervals/cells for $(x_1, ..., x_L)$.

Unconstrained model (2)

- This corresponds to a random classification of data by using rectangles/or boxes (= Cartesian products of intervals on the covariate axes).

- Assuming a constant hazard rate within each cell j, the log-likelihood contribution for observation $i = 1, \ldots, n$ is of the Poisson form
$$d_{ij}(\theta_j + \beta' z_i) - y_{ij} \exp(\theta_j + \beta' z_i),$$
where
  - d_{ij} is event indicator, and
  - y_{ij} is follow-up time in cell j.

Motivation for model

- The random change point model applied here is motivated by
  - its computational simplicity
  - its straightforward extendibility to the multivariate case
  - the relative robustness of numerical probability evaluations such as predictive distributions (generally involving an integration of densities/intensities) w.r.t. to the smoothness properties of the functions involved in such models, and
  - the fact that, if needed, model averaging over non-smooth models will generally produce smooth reconstructions of the functional relationships considered (e.g., in terms of the pointwise posterior medians or means).

Prior for change points

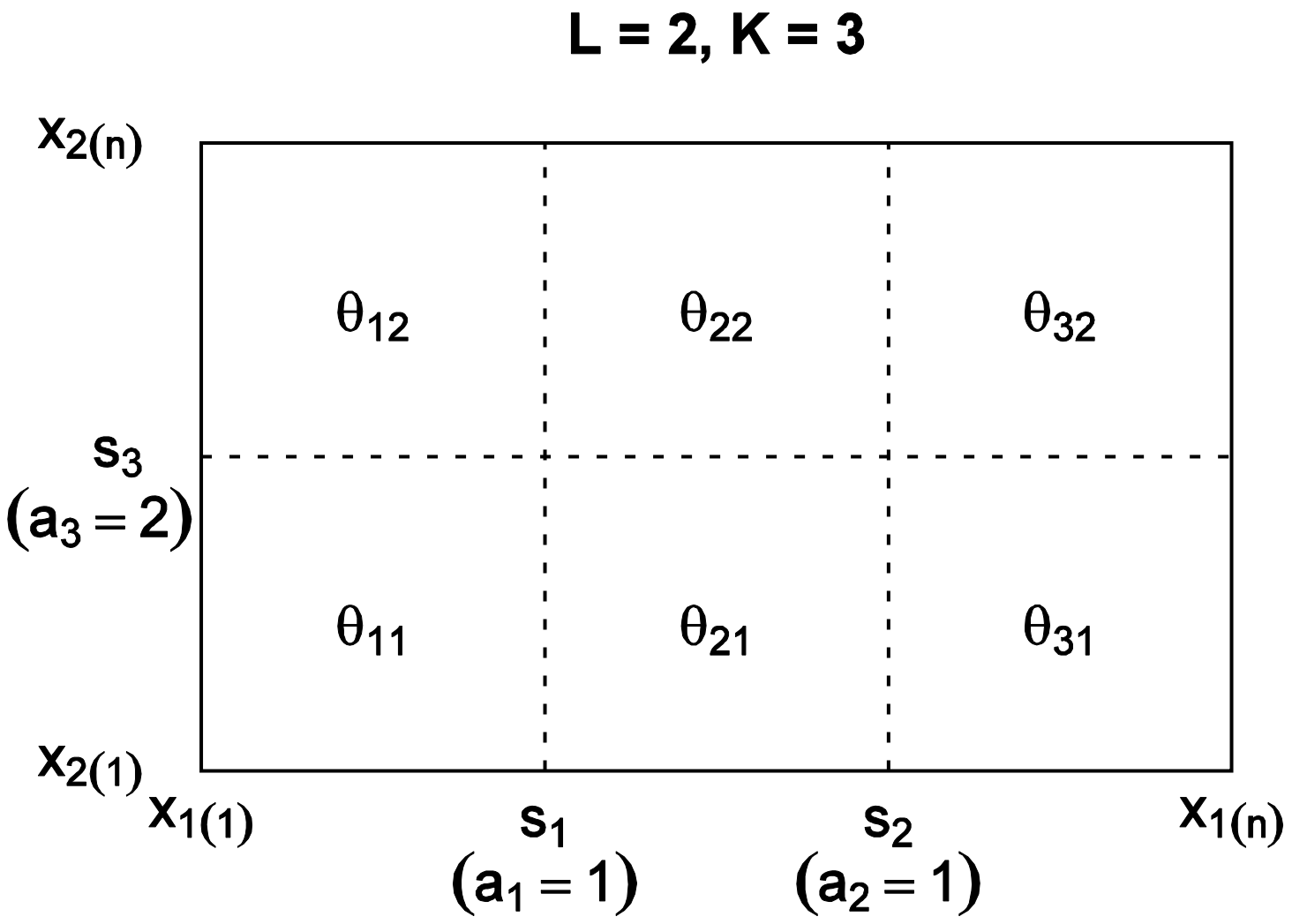
- A common pool of change points is allocated between the covariate axes; they are generally utilised where they produce the best fit to data.
- Distribution $p(K)$, $K = 0, \ldots, K_{MAX}$ for the total number of change points can be used to control the smoothness of the function estimate.
- Each change point k, $k = 1, \ldots, K$, is allocated to one of the L covariate axes with prior probability $p(a_k) = 1/L$, with $a_k = 1, \ldots, L$.

Prior for change points (2)

- Prior for change point locations $p(s_1, ..., s_K \mid a_1, ..., a_K, K)$?

- Given total number and allocation, locations of K_l points allocated for axis l can be distributed as
  - order statistics of K_l points uniformly distributed on $[x_{l(1)}, x_{l(n)}]$, or
  - even-numbered order statistics of $2K_l + 1$ points uniformly distributed on $[x_{l(1)}, x_{l(n)}]$. This alternative places the step positions further away from each other, avoiding too many short steps (Green, 1995).

Illustration



MCMC algorithm

- Four types of moves
  1. Propose to add a new, or remove an existing change point and propose new values for related θ parameters.
  2. Propose to change location of an existing change point and propose new values for related θ parameters.
  3. Update the values of parameters θ
  4. Update the values of parameters β

MCMC algorithm (2)

- Proposal for parameter θ_j in a class j is
  $N(\log(\sum_i d_{ij} / \sum_i y_{ij} \exp(\beta' z_i)), 1/ \sum_i d_{ij})$,
  which corresponds to the maximum likelihood estimator for log rate parameter.
- If $\sum_i d_{ij} = 0$, θ_j is not defined.
- This proposal is independent of the current state of the chain; no Jacobian term needed in the Metropolis-Hastings acceptance ratio (cf. Dellaportas et al., 2002).

Monotonic restrictions

- Back to the example with $L = 2$ and $K = 3$; Assuming isotonic effect for both axes, resulting partial ordering:

	θ_{11}	θ_{21}	θ_{31}	θ_{12}	θ_{22}	θ_{32}
θ_{11}	$=$	$\leq$	$\leq$	$\leq$	$\leq$	$\leq$
θ_{21}	$\geq$	$=$	$\leq$		$\leq$	$\leq$
θ_{31}	$\geq$	$\geq$	$=$			$\leq$
θ_{12}	$\geq$			$=$	$\leq$	$\leq$
θ_{22}	$\geq$	$\geq$		$\geq$	$=$	$\leq$
θ_{32}	$\geq$	$\geq$	$\geq$	$\geq$	$\geq$	$=$

How to apply monotonicity?

1. Sample from the unconstrained model and extract monotonic part from the resulting posterior sample (Holmes & Heard, 2003).
2. Use the unconstrained model as a proposal and incorporate monotonic constraints in the prior for θ. Moves which violate monotonicity are rejected.
3. Construct a monotonic proposal.

Applying monotonicity (2)

- Alternatives 1 and 2 are easy to implement because sampling from the unconstrained model is computationally straightforward.
- However, it is less clear whether use of the unconstrained model will generally result in good mixing in the space of monotonic submodels.
- It is not obvious how to construct good monotonic proposal distributions, particularly in higher dimensions.

Applying monotonicity (3)

- The number of θ parameters in our ”piecewise constant” model is
  $$(K_1 + 1) \ldots (K_L + 1),$$
  where K_l is the number of change points allocated on axis l.
- For example, when a new change point is proposed on axis 1, the number of proposed new θ parameters needed for such a move is
  $$2(K_2 + 1) \ldots (K_L + 1).$$

Applying monotonicity (4)

- A good monotonic proposal distribution would be such that acceptance probability for birth/death moves for change points does not suffer when L (= # covariates in θ) increases.

- Further work is needed here.

Question

- Is this BNR?

Illustration

- A cohort of 4369 men aged 54-77 from southern and western Finland examined in 1992-1993 and free of CVD at the time of examination.
- Prospective follow-up for CVD events (coronary heart disease + ischaemic stroke) to the end of year 1999 .
- 568 first CVD events were recorded during the follow-up.

Illustration (2)

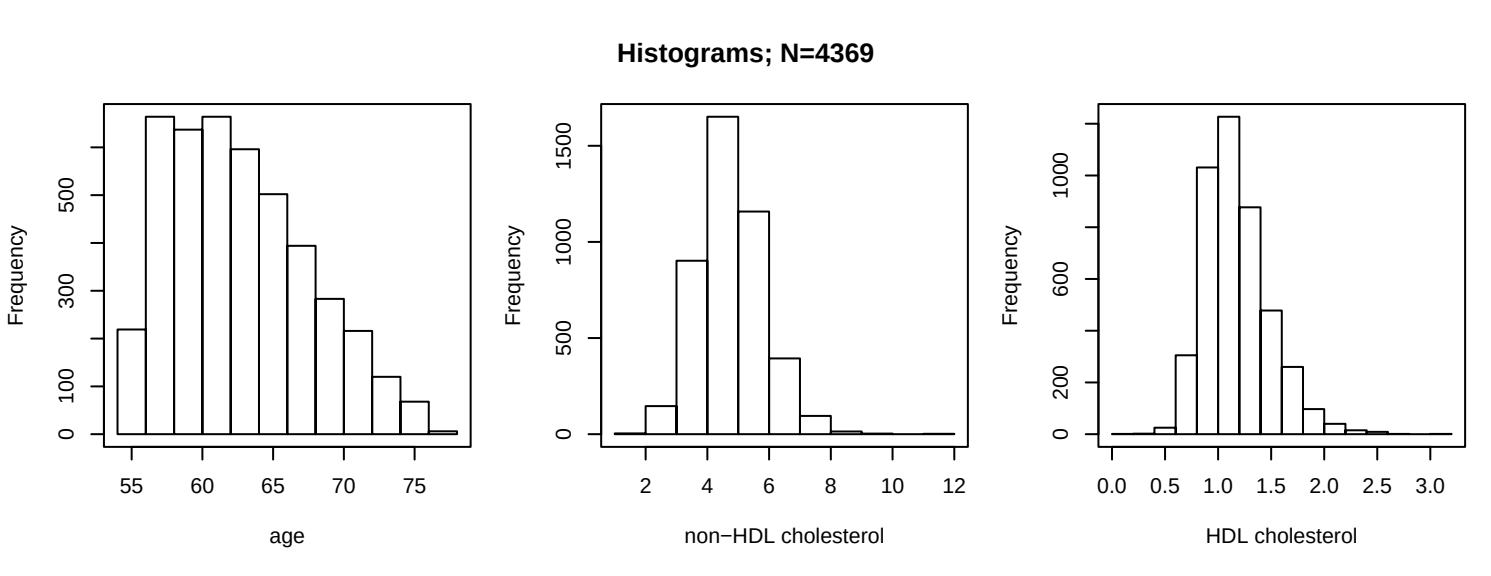
- Consider here three risk factors; age and non-HDL cholesterol (isotonic effect assumed for CVD risk) and HDL cholesterol (antitonic effect assumed for CVD risk).
- Is there evidence of nonlinearity/non-additivity when the factors are considered pairwise with the nonparametric model?

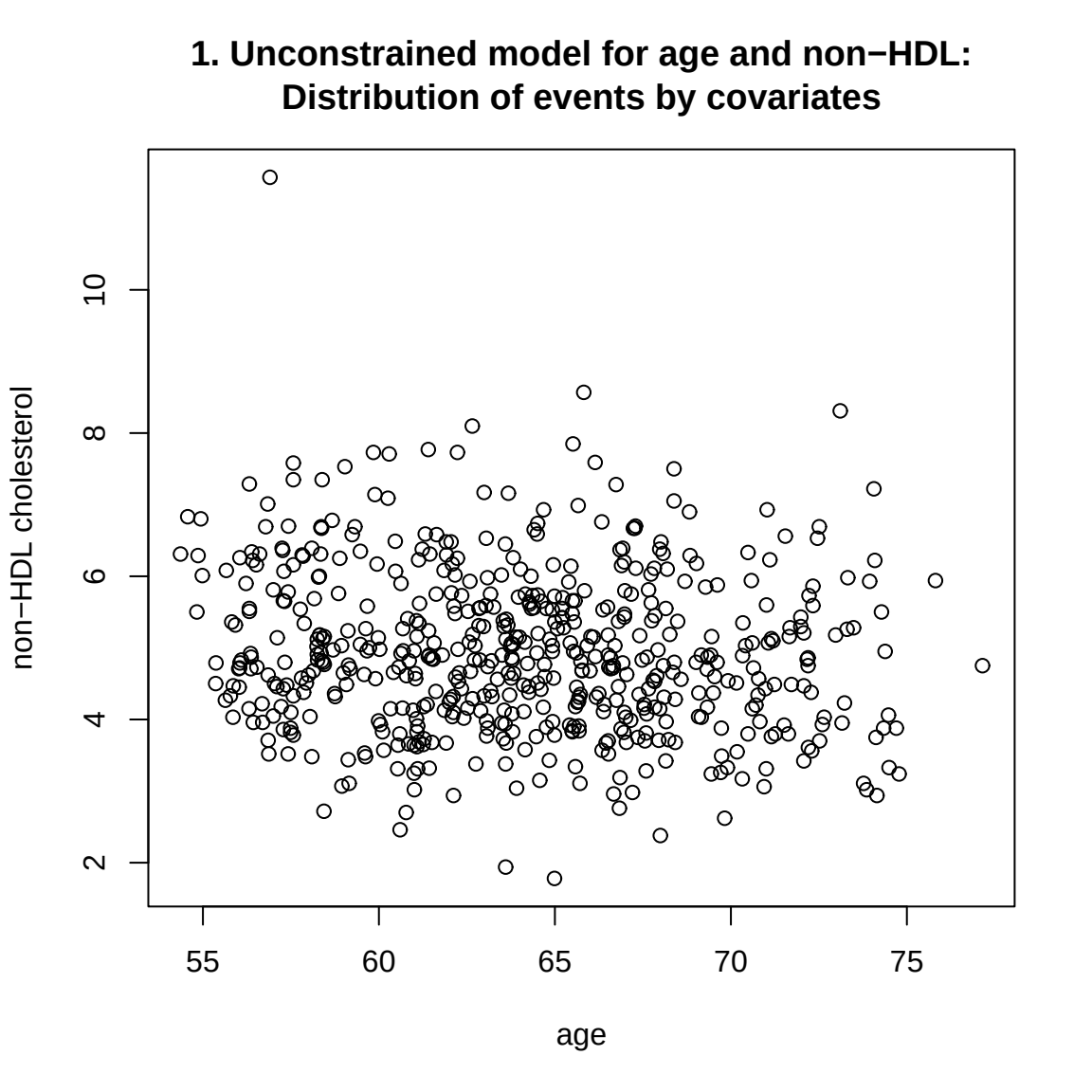
Different models

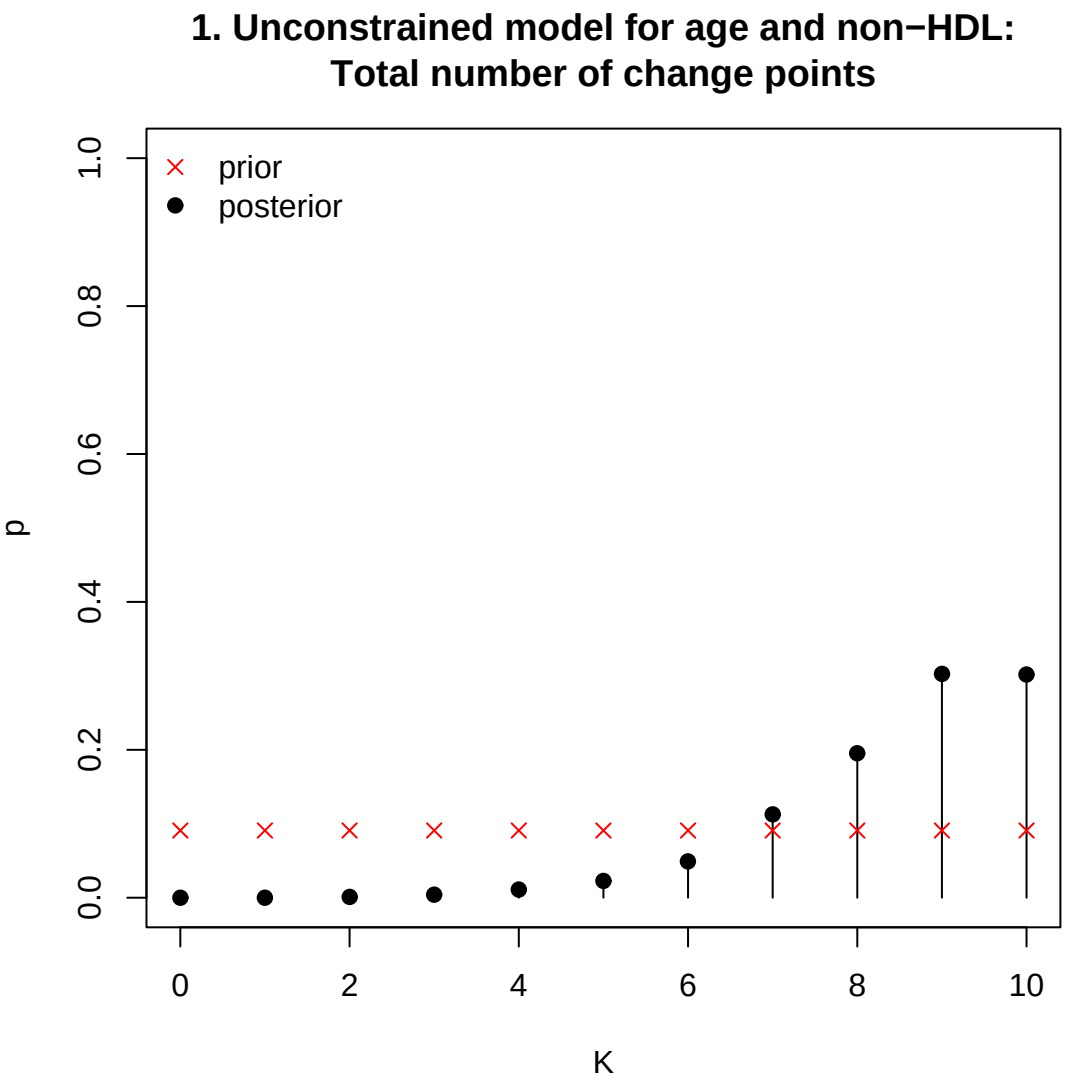
1. Unconstrained model for age and non-HDL
2. Unconstrained model for ranks of age and non-HDL
3. Constrained model for age and non-HDL
4. Constrained model for ranks of age and non-HDL

5. Unconstrained model for age and HDL
6. Unconstrained model for ranks of age and HDL
7. Constrained model for age and HDL
8. Constrained model for ranks of age and HDL

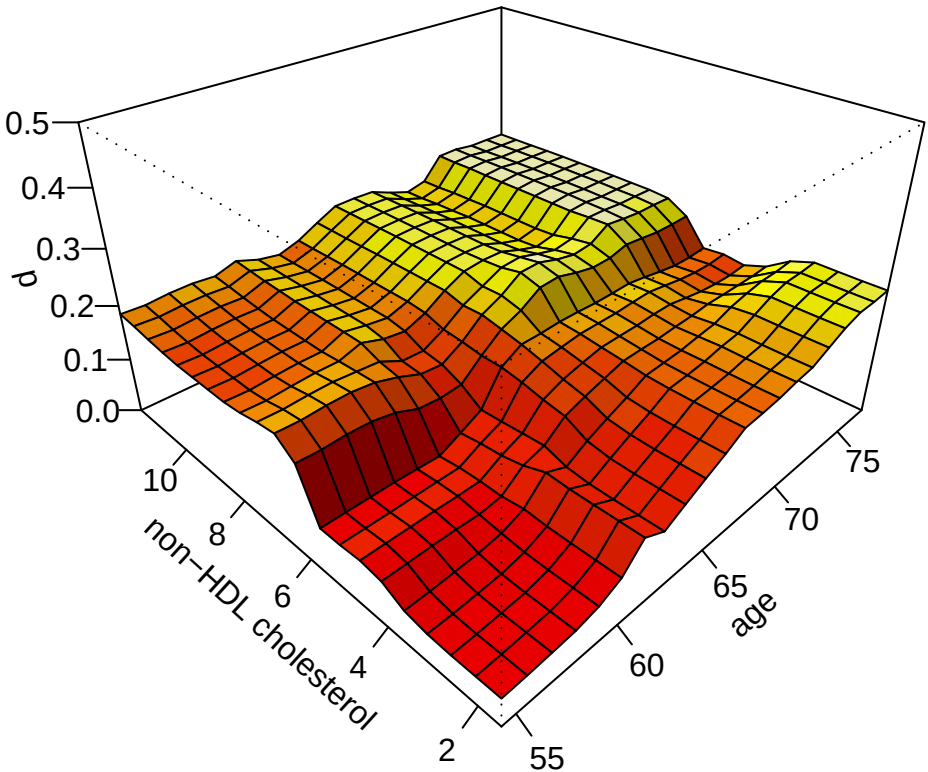
9. Unconstrained model for non-HDL and HDL
10. Unconstrained model for ranks of non-HDL and HDL
11. Constrained model for non-HDL and HDL
12. Constrained model for ranks of non-HDL and HDL


Histograms; N=4369
Frequency
0
100
300
500
55
60
65
70
75
age
Frequency
0
500
1000
1500
2
4
6
8
10
12
non−HDL cholesterol
Frequency
0
200
600
1000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
HDL cholesterol

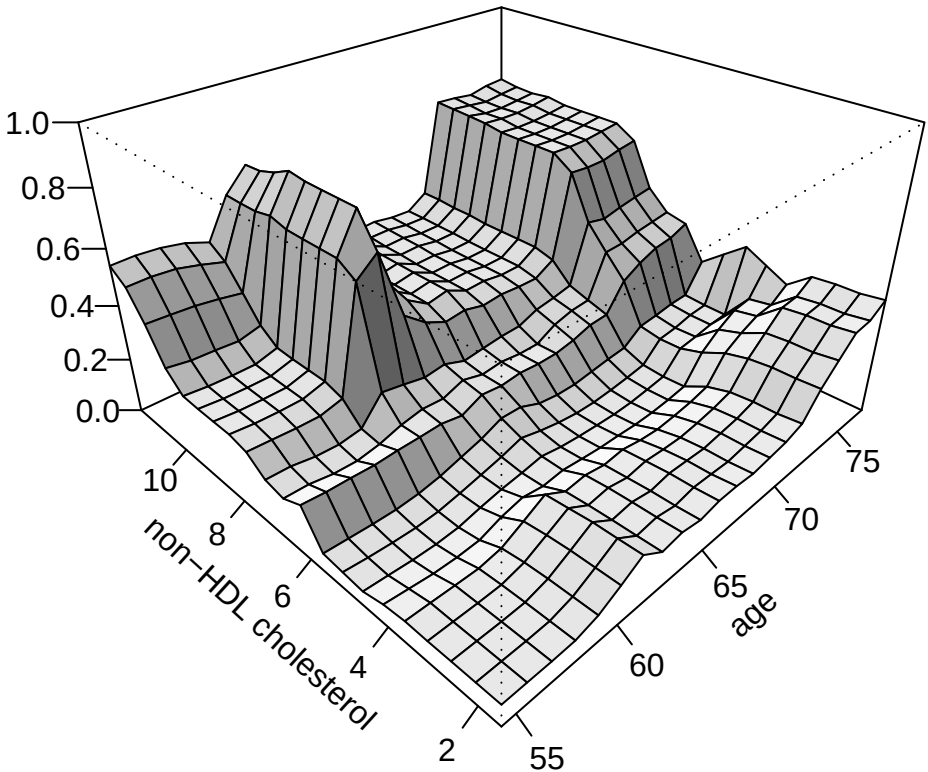

1. Unconstrained model for age and non−HDL:
Distribution of events by covariates
non−HDL cholesterol
2
4
6
8
10
55
60
65
70
75
age


1. Unconstrained model for age and non−HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

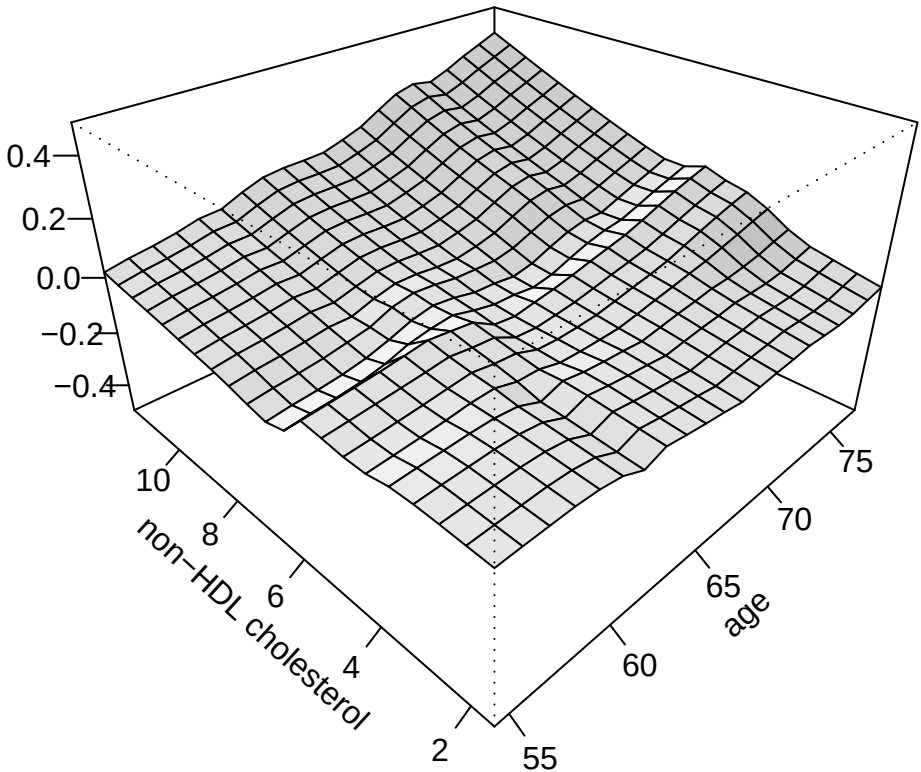
1. Unconstrained model for age and non−HDL: Posterior median 7 year risk

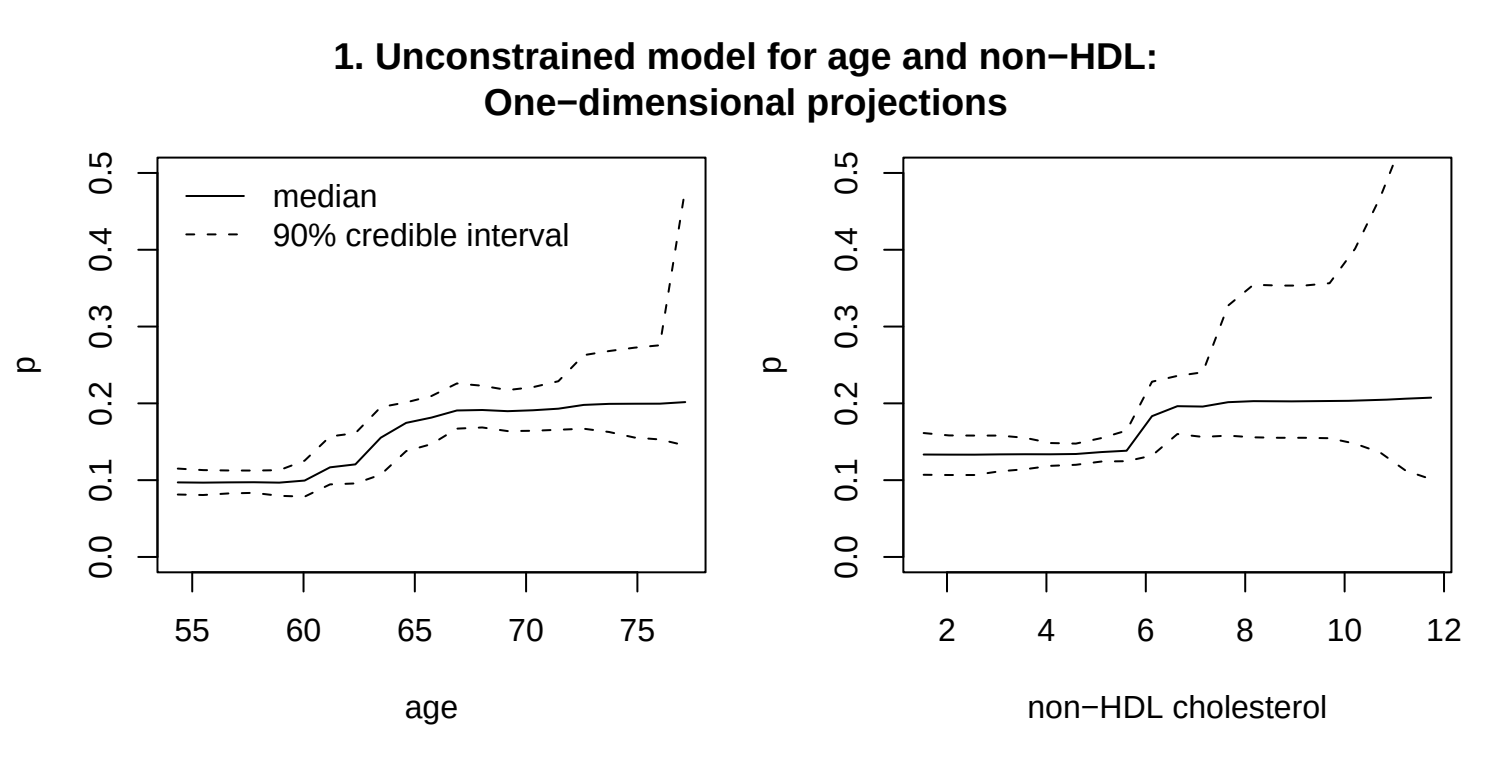


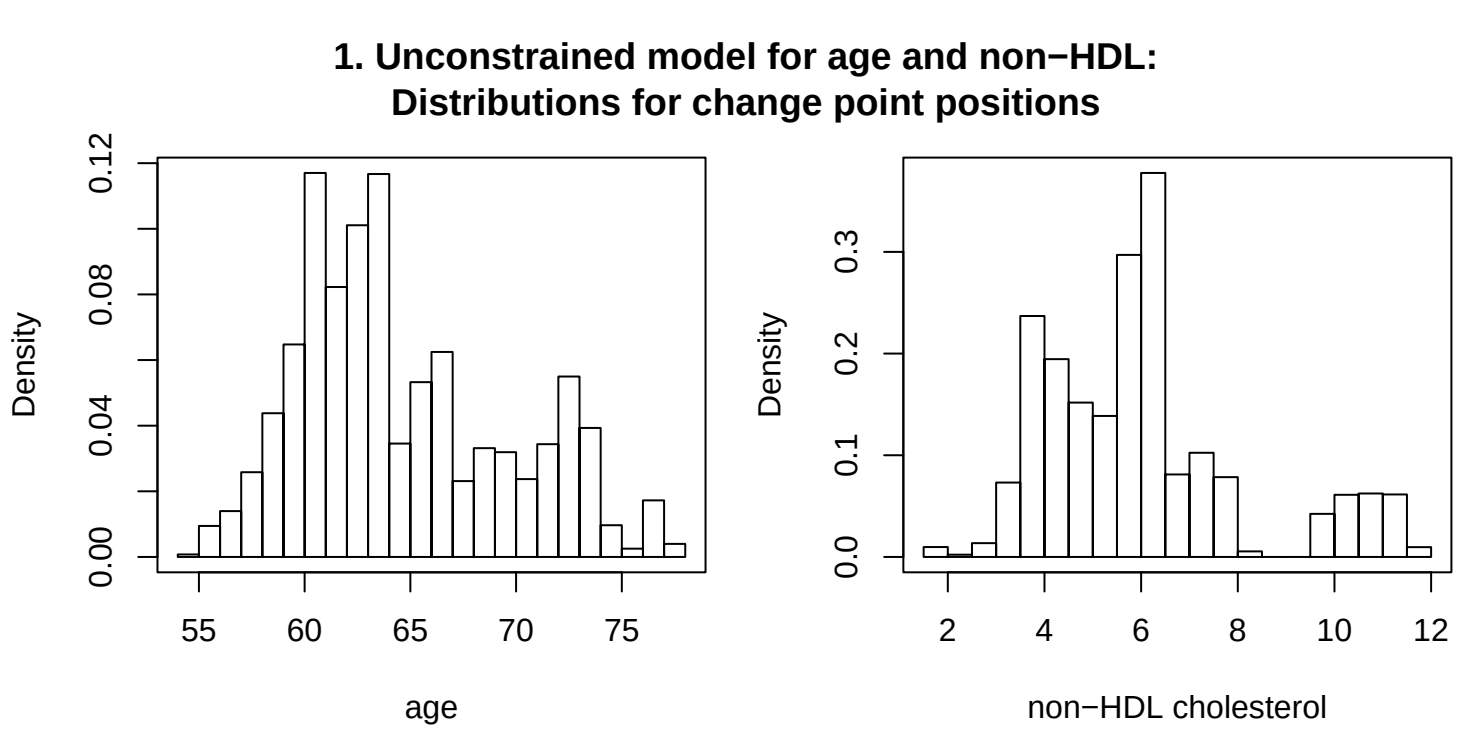
1. Unconstrained model for age and non−HDL: Length of 90% credible interval

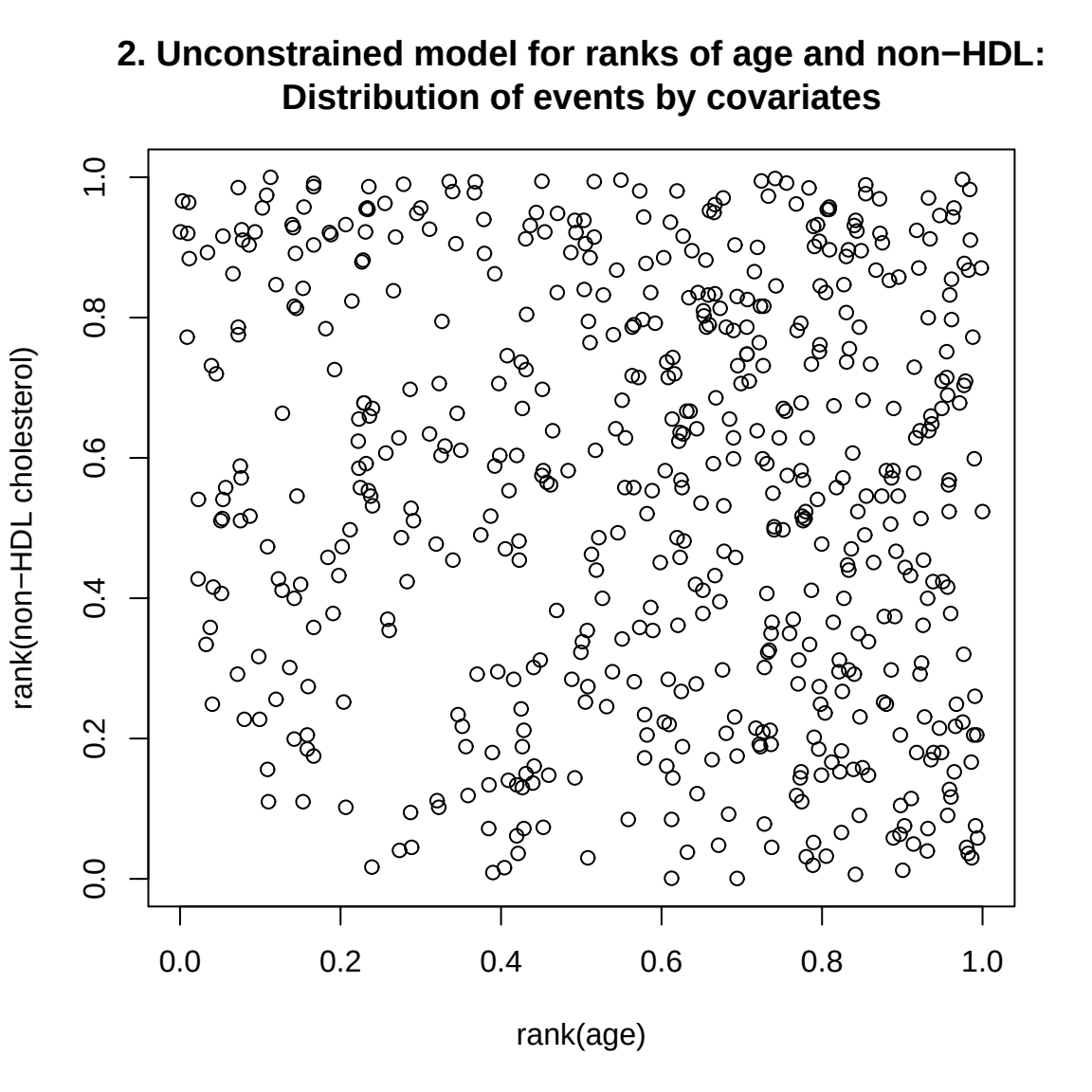


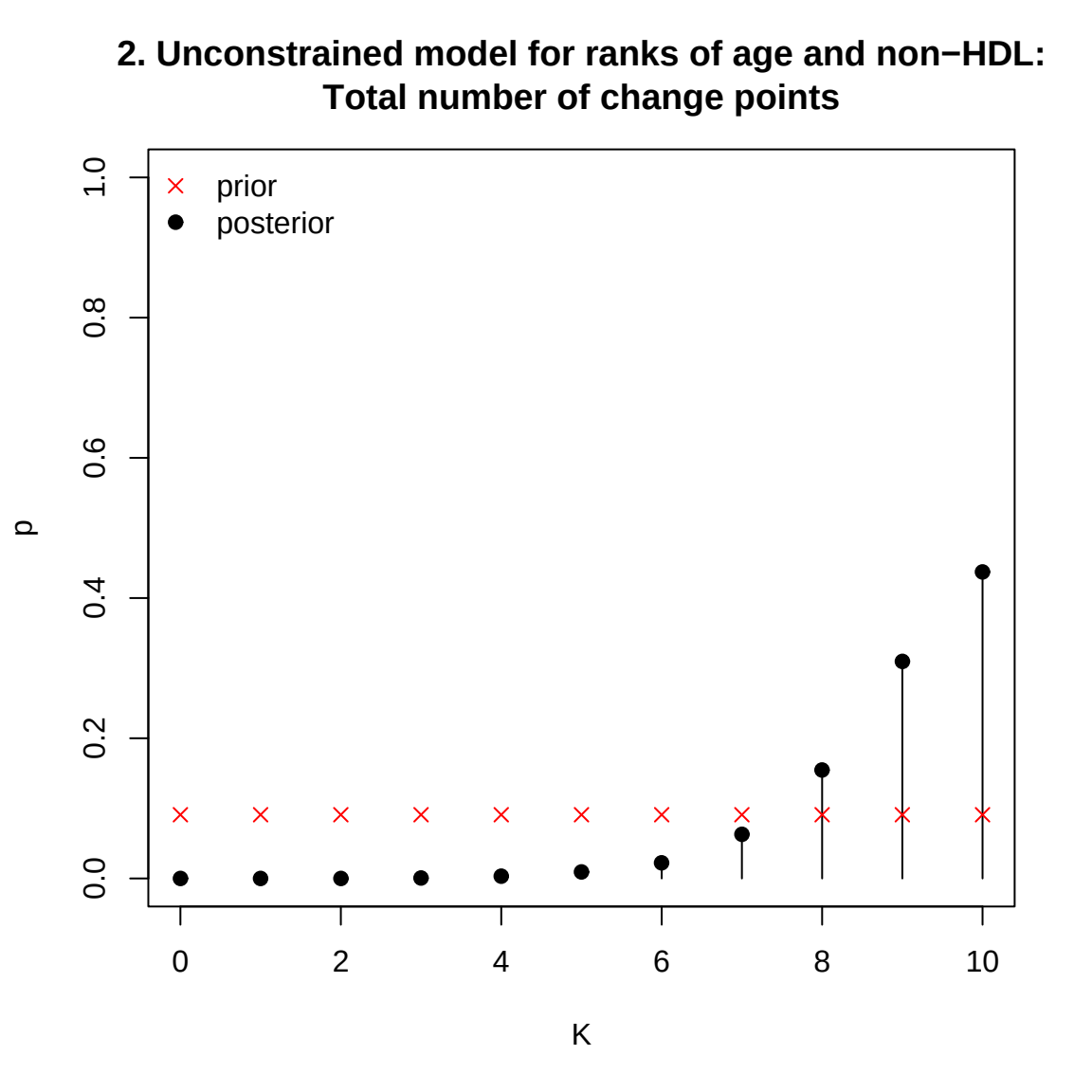
1. Unconstrained model for age and non−HDL: Difference to proportional hazards model



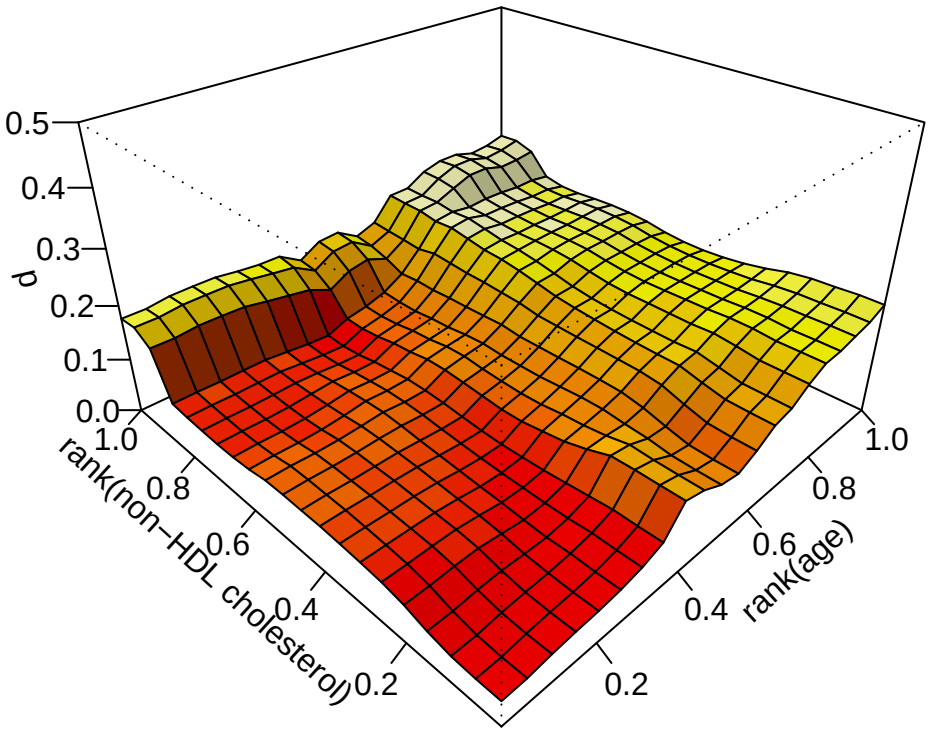

1. Unconstrained model for age and non−HDL:
One−dimensional projections
median
90% credible interval
p
0.0
0.1
0.2
0.3
0.4
0.5
55
60
65
70
75
age
2
4
6
8
10
12
non−HDL cholesterol


1. Unconstrained model for age and non−HDL:
Distributions for change point positions
Density
0.00
0.04
0.08
0.12
55
60
65
70
75
age
Density
0.0
0.1
0.2
0.3
2
4
6
8
10
12
non−HDL cholesterol

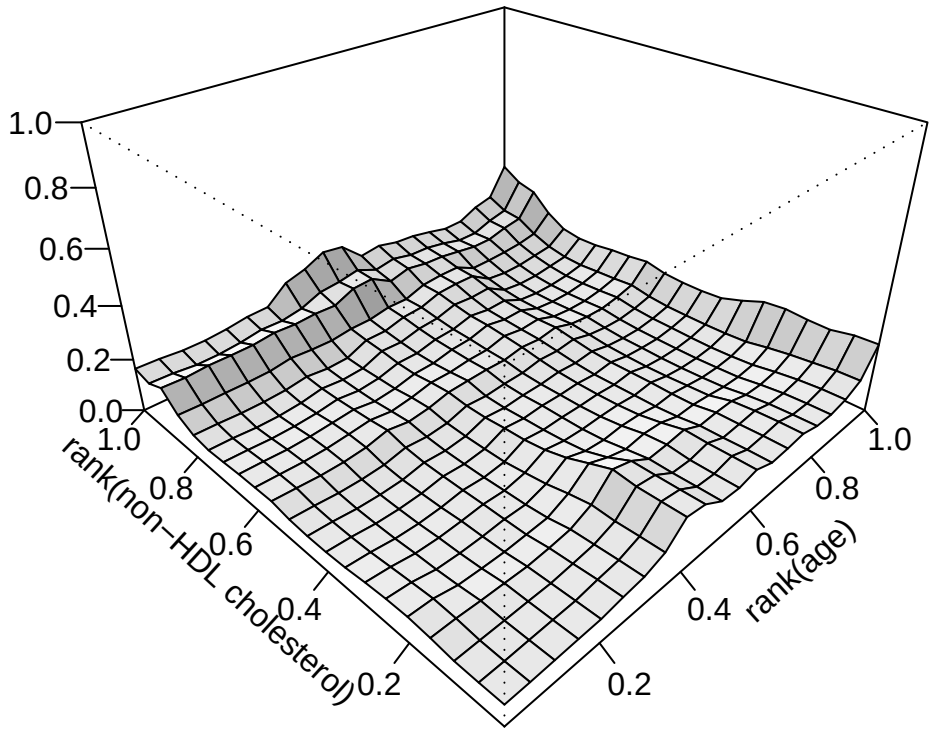

2. Unconstrained model for ranks of age and non−HDL:
Distribution of events by covariates
rank(non−HDL cholesterol)
0.0
0.2
0.4
0.6
0.8
1.0
rank(age)
0.0
0.2
0.4
0.6
0.8
1.0


2. Unconstrained model for ranks of age and non−HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

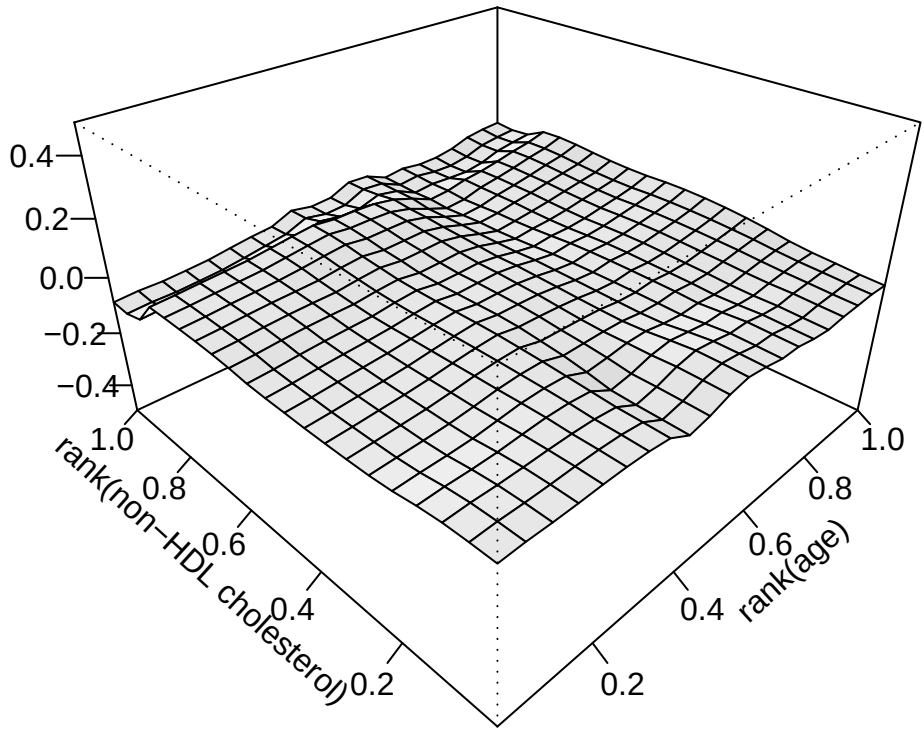
2. Unconstrained model for ranks of age and non−HDL: Posterior median 7 year risk

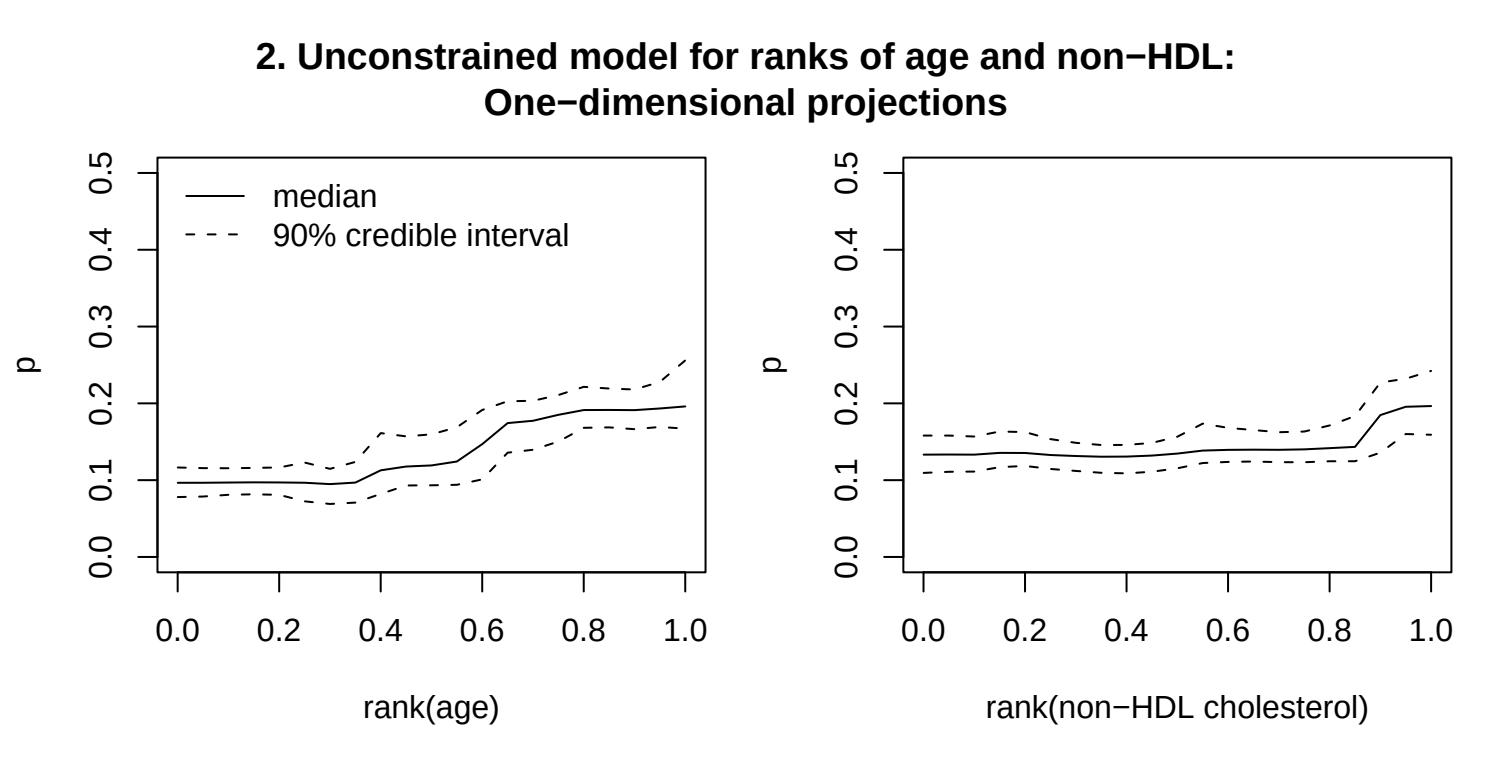


2. Unconstrained model for ranks of age and non−HDL: Length of 90% credible interval

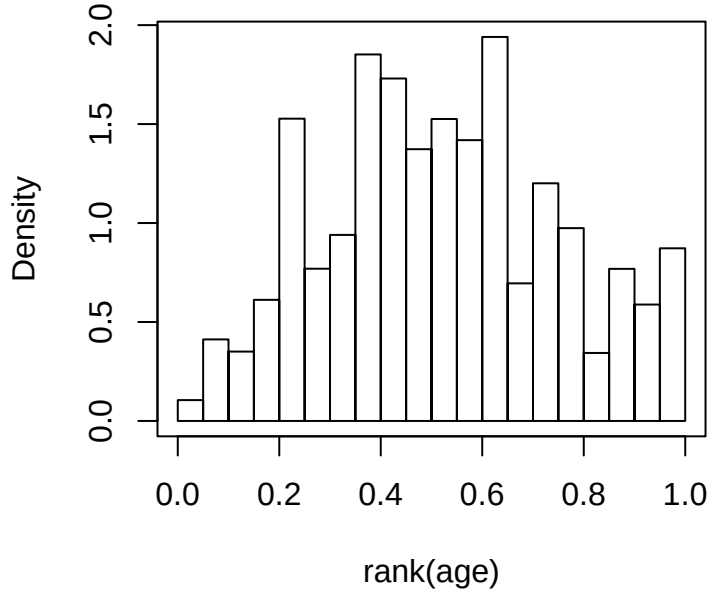


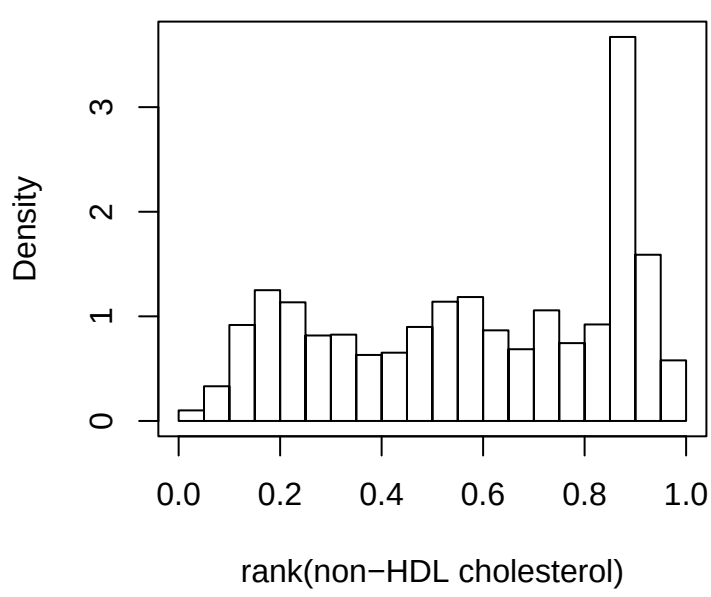
2. Unconstrained model for ranks of age and non−HDL: Difference to proportional hazards model

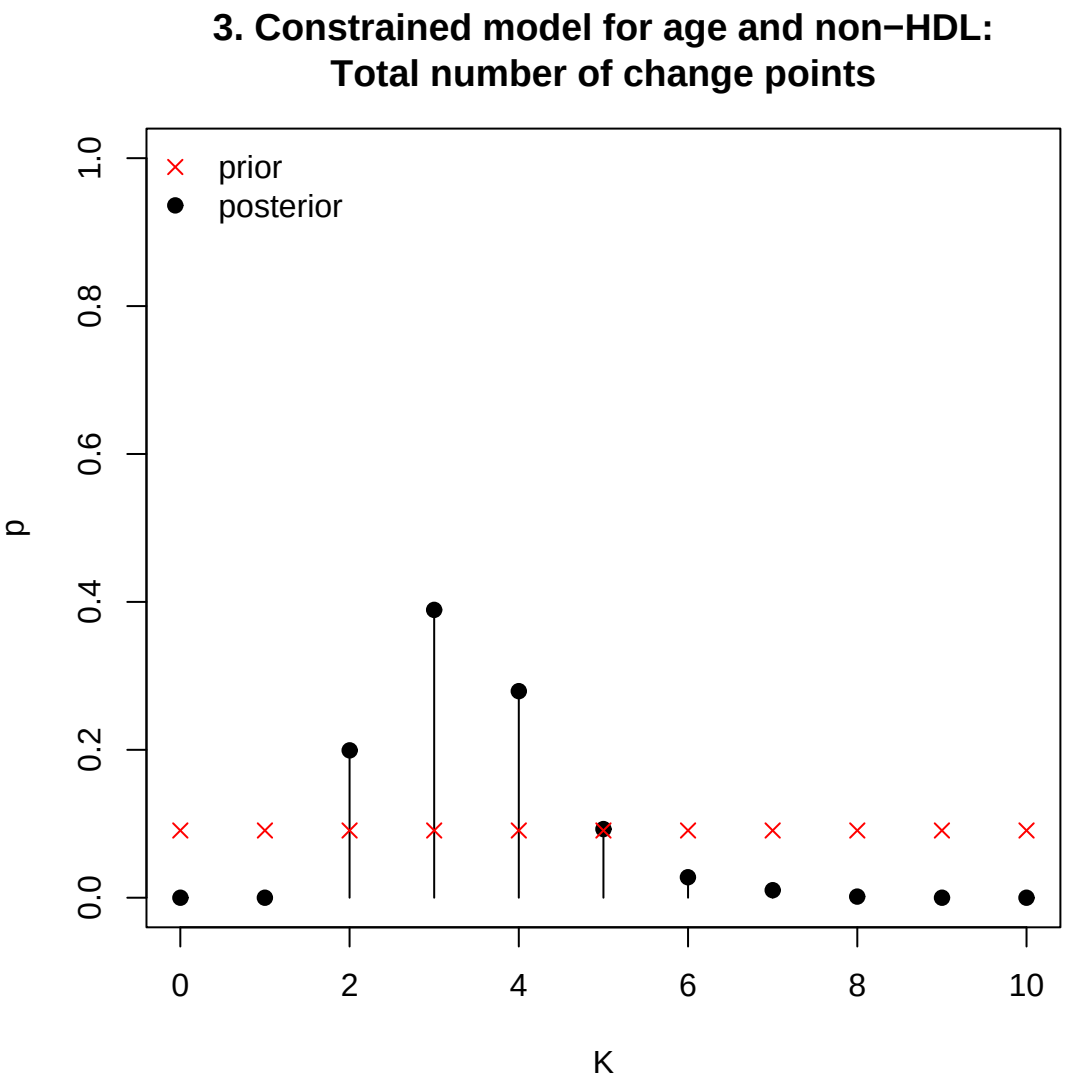



2. Unconstrained model for ranks of age and non−HDL:
One−dimensional projections
median
90% credible interval
p
0.0
0.1
0.2
0.3
0.4
0.5
0.0
0.2
0.4
0.6
0.8
1.0
rank(age)
rank(non−HDL cholesterol)

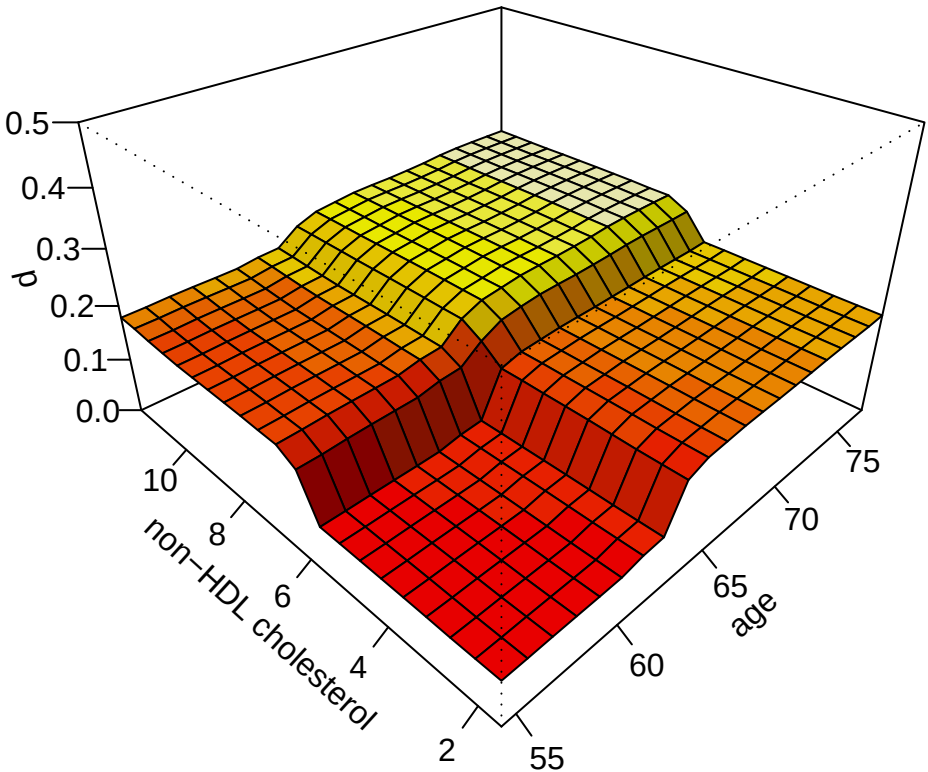
2. Unconstrained model for ranks of age and non−HDL: Distributions for change point positions



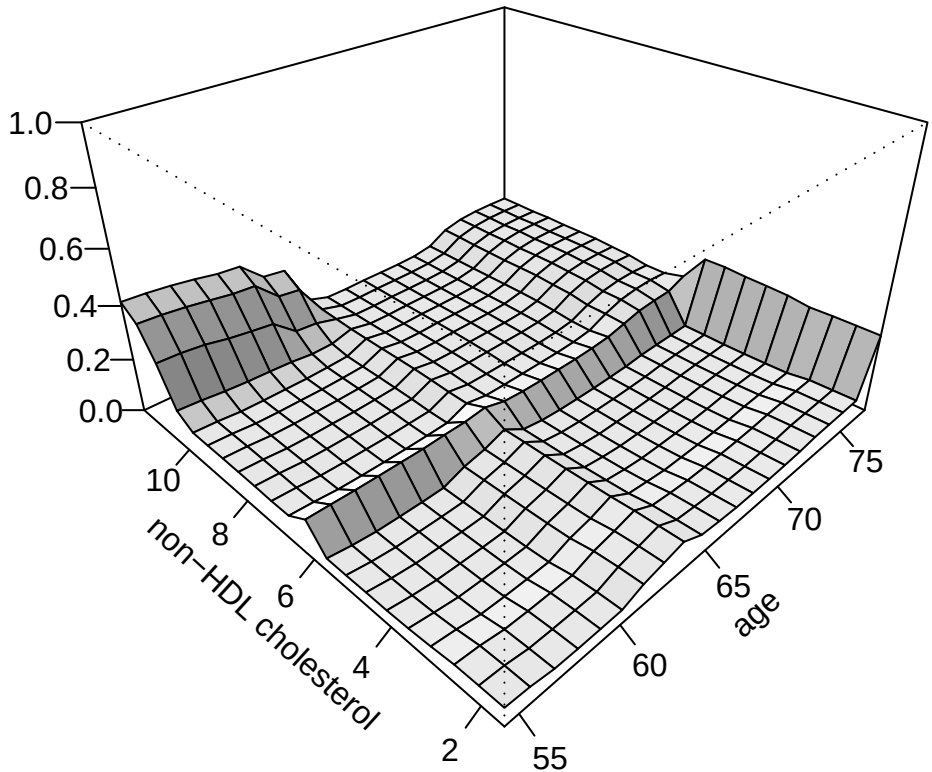



3. Constrained model for age and non−HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

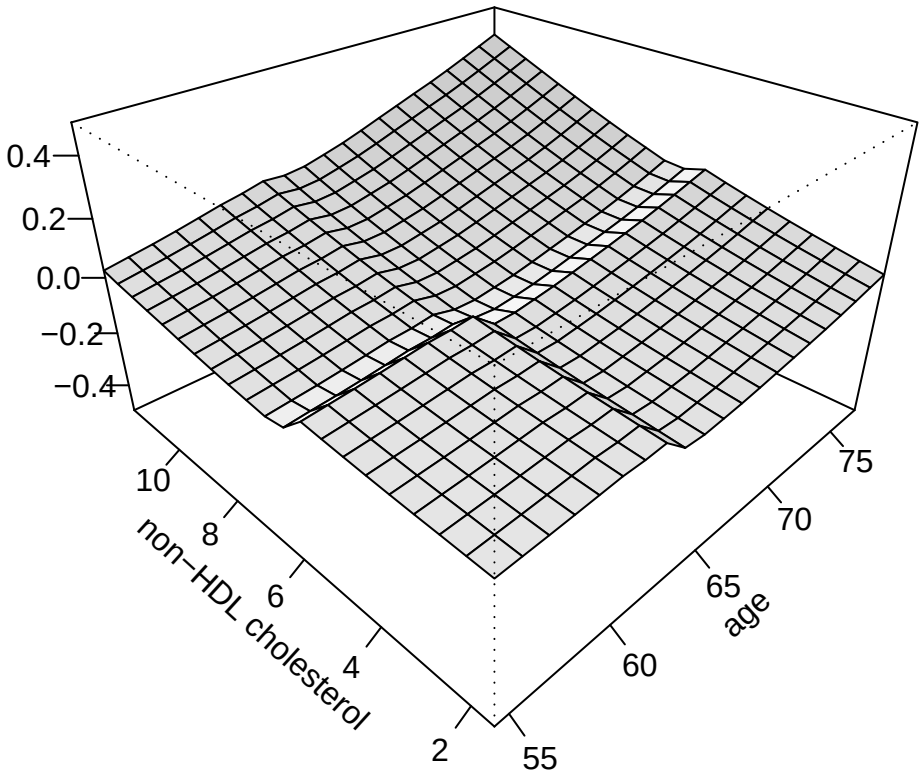
3. Constrained model for age and non−HDL: Posterior median 7 year risk

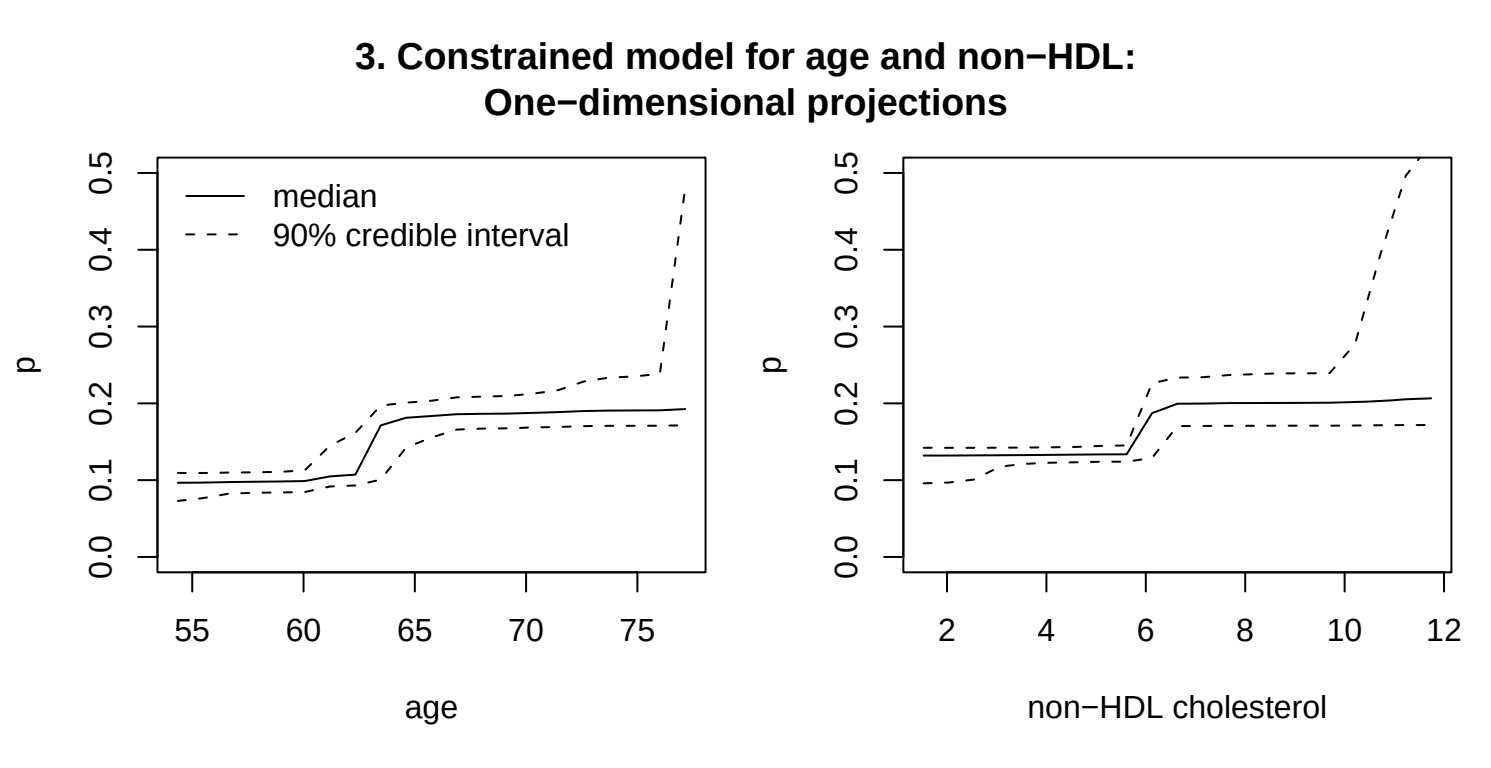


3. Constrained model for age and non−HDL: Length of 90% credible interval

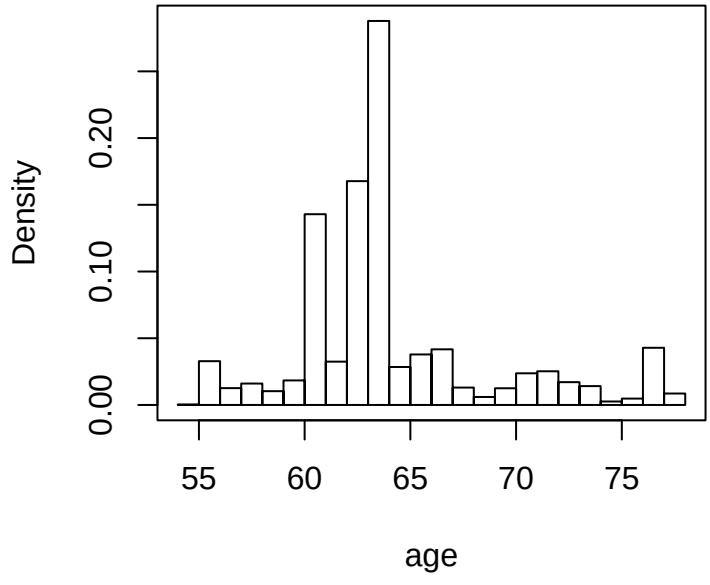


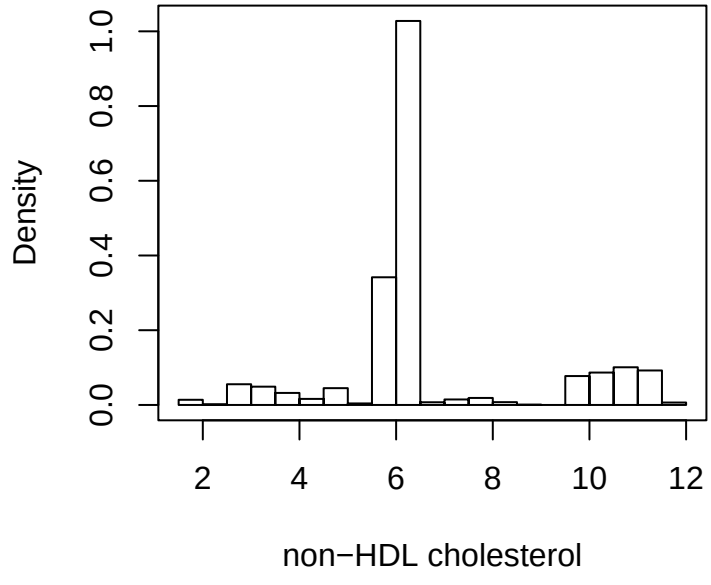
3. Constrained model for age and non−HDL: Difference to proportional hazards model

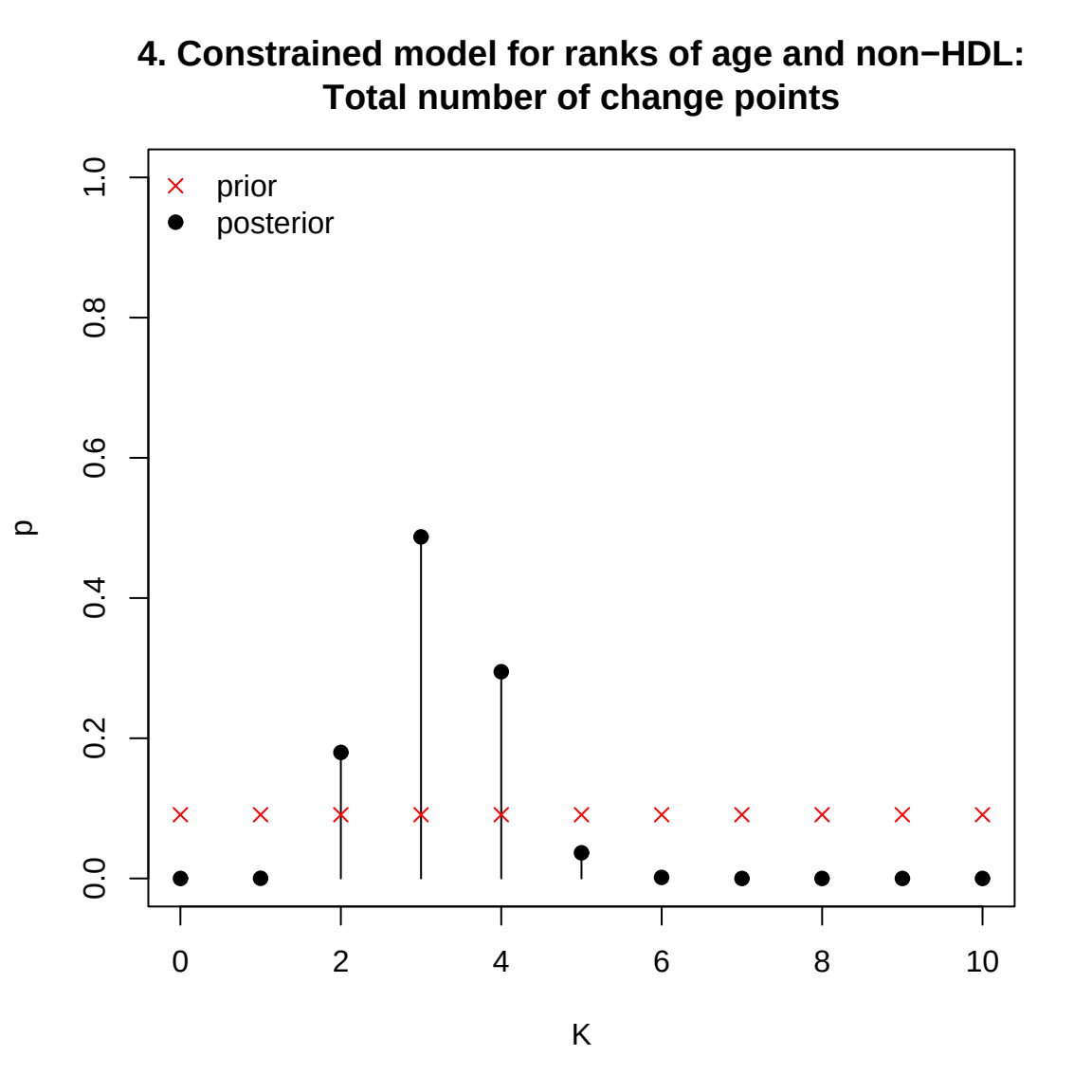



3. Constrained model for age and non−HDL:
One−dimensional projections
median
90% credible interval
p
0.0
0.1
0.2
0.3
0.4
0.5
55
60
65
70
75
age
2
4
6
8
10
12
non−HDL cholesterol

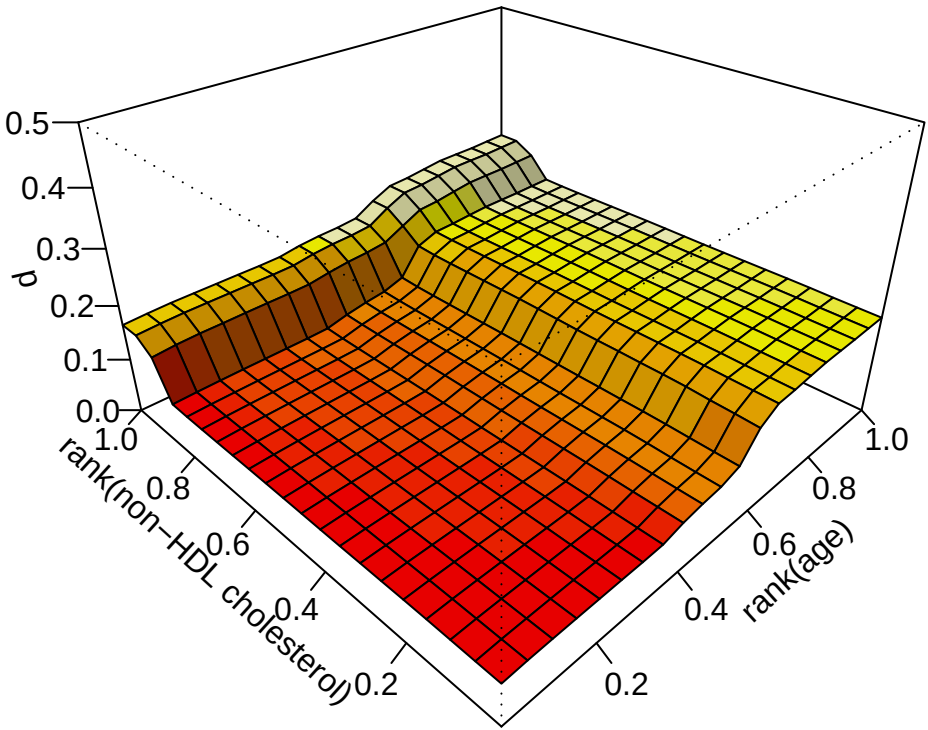
3. Constrained model for age and non−HDL: Distributions for change point positions



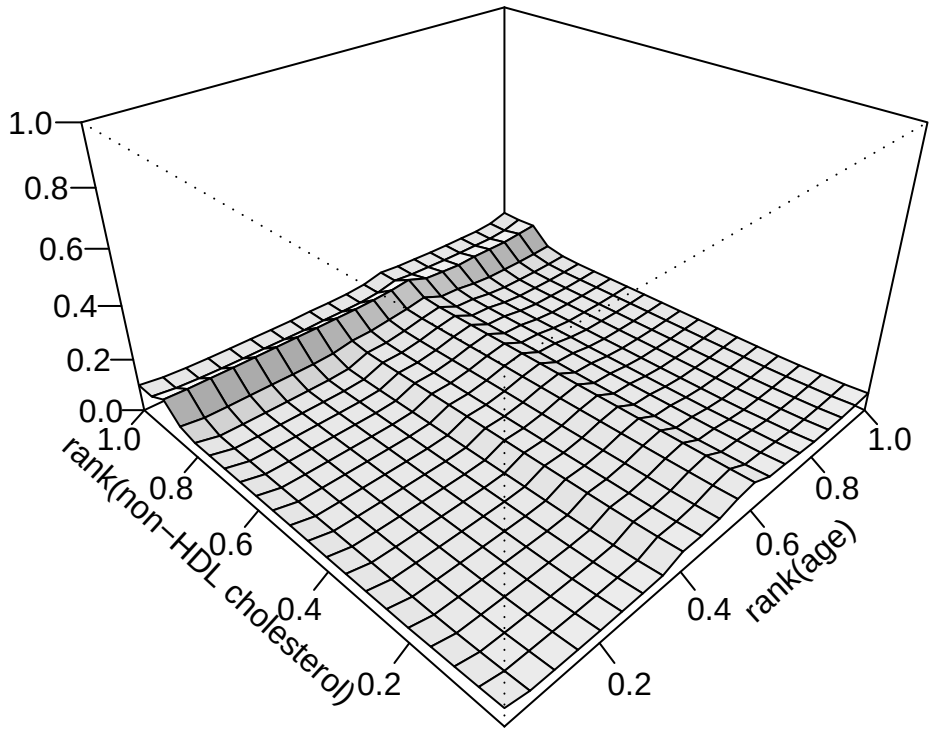



4. Constrained model for ranks of age and non−HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

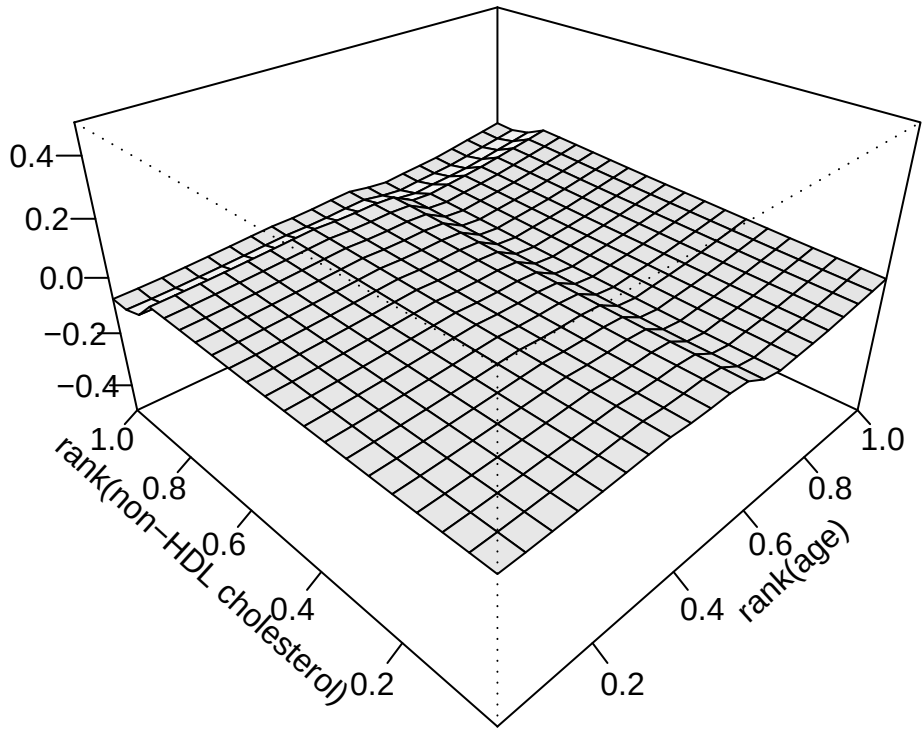
4. Constrained model for ranks of age and non−HDL: Posterior median 7 year risk



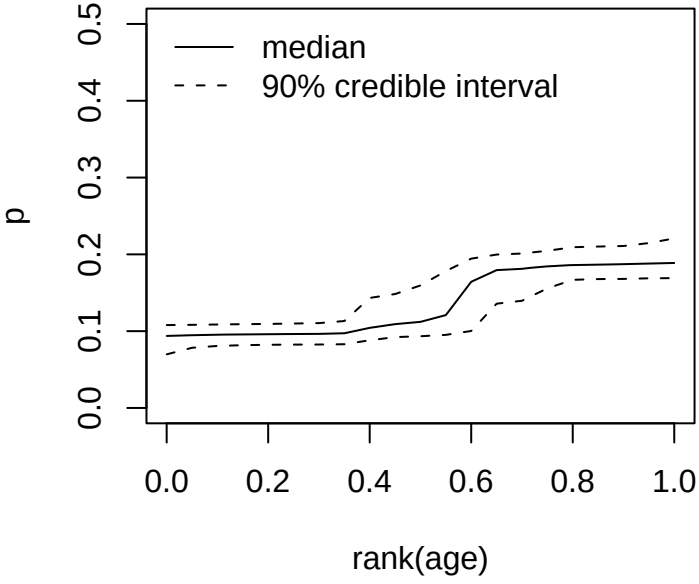
4. Constrained model for ranks of age and non−HDL: Length of 90% credible interval

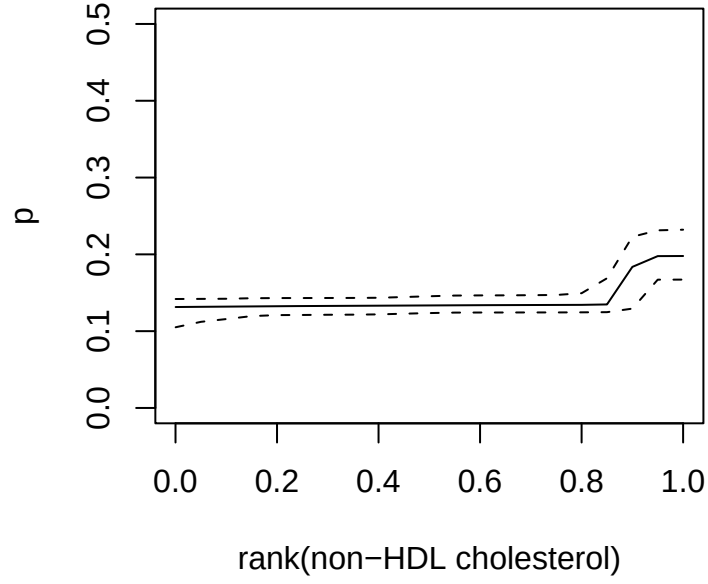


4. Constrained model for ranks of age and non−HDL: Difference to proportional hazards model

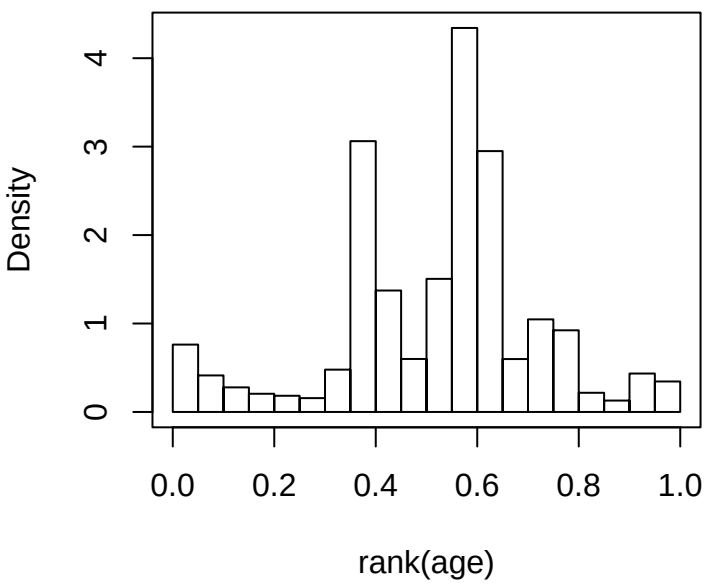


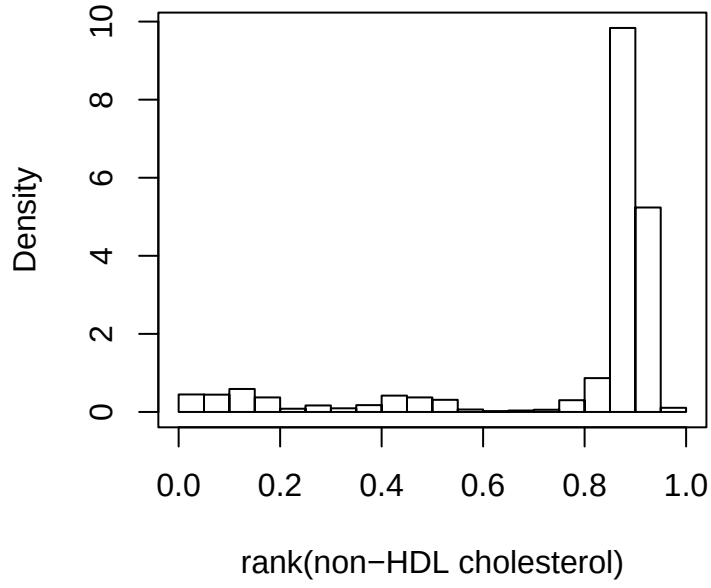
4. Constrained model for ranks of age and non−HDL: One−dimensional projections



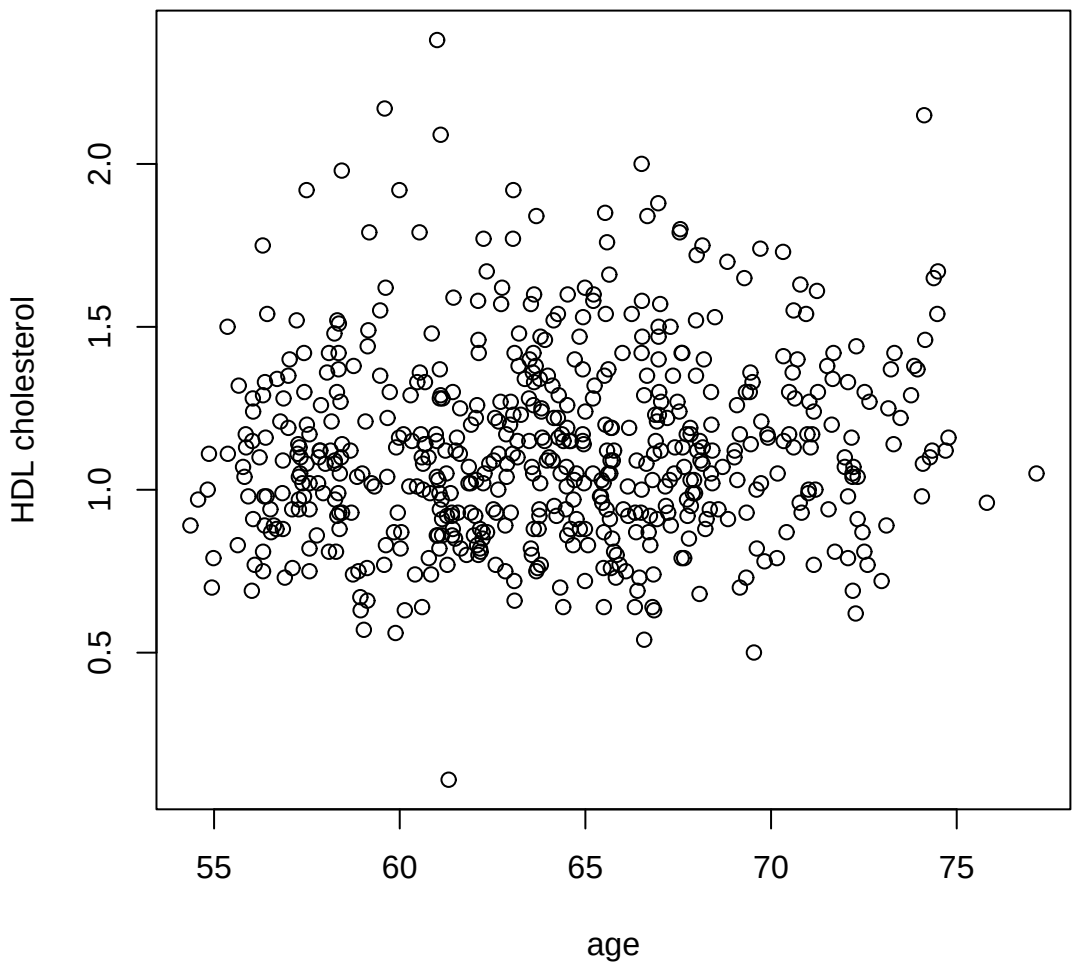


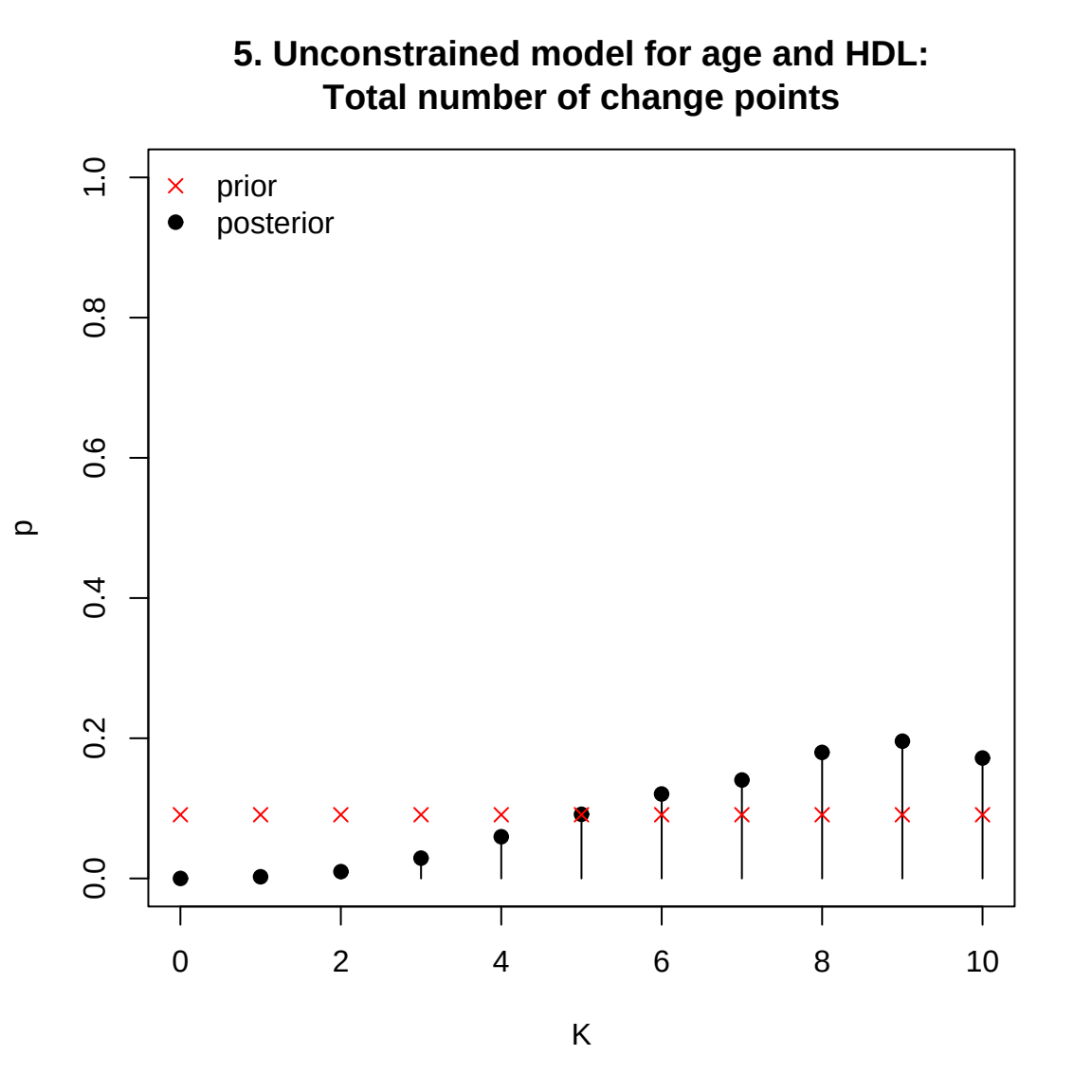
4. Constrained model for ranks of age and non−HDL: Distributions for change point positions



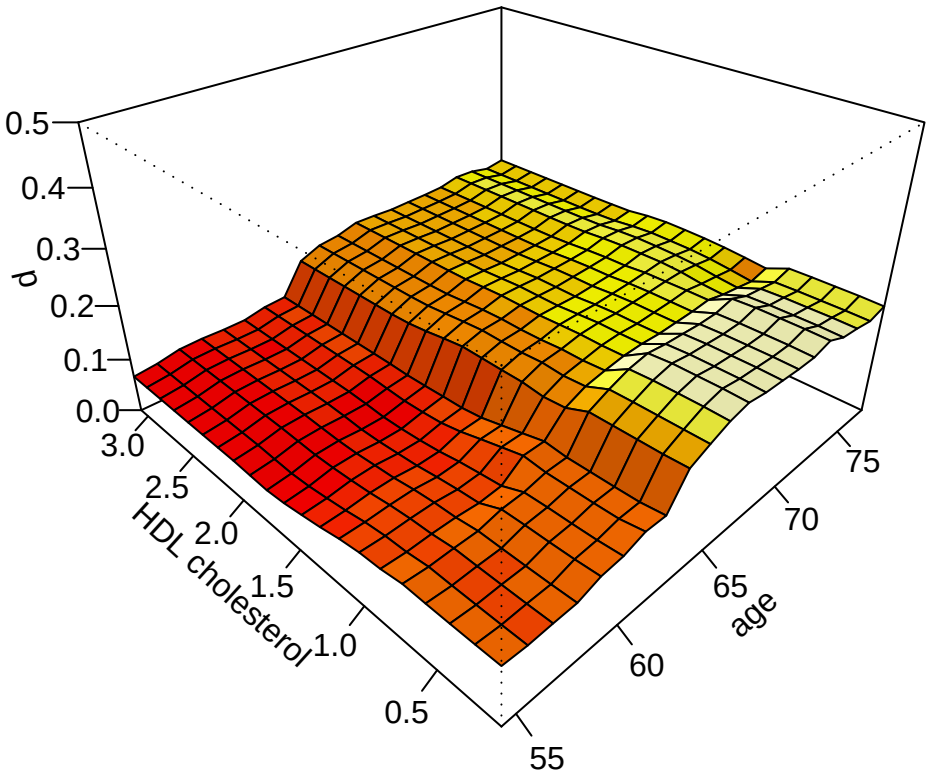


5. Unconstrained model for age and HDL: Distribution of events by covariates

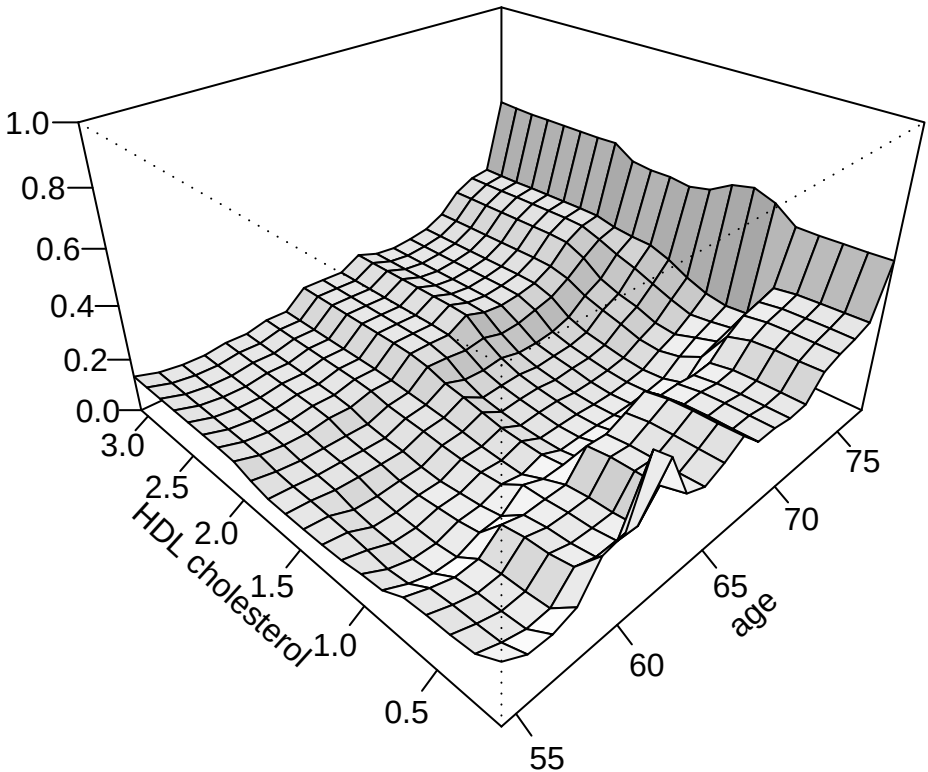



5. Unconstrained model for age and HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

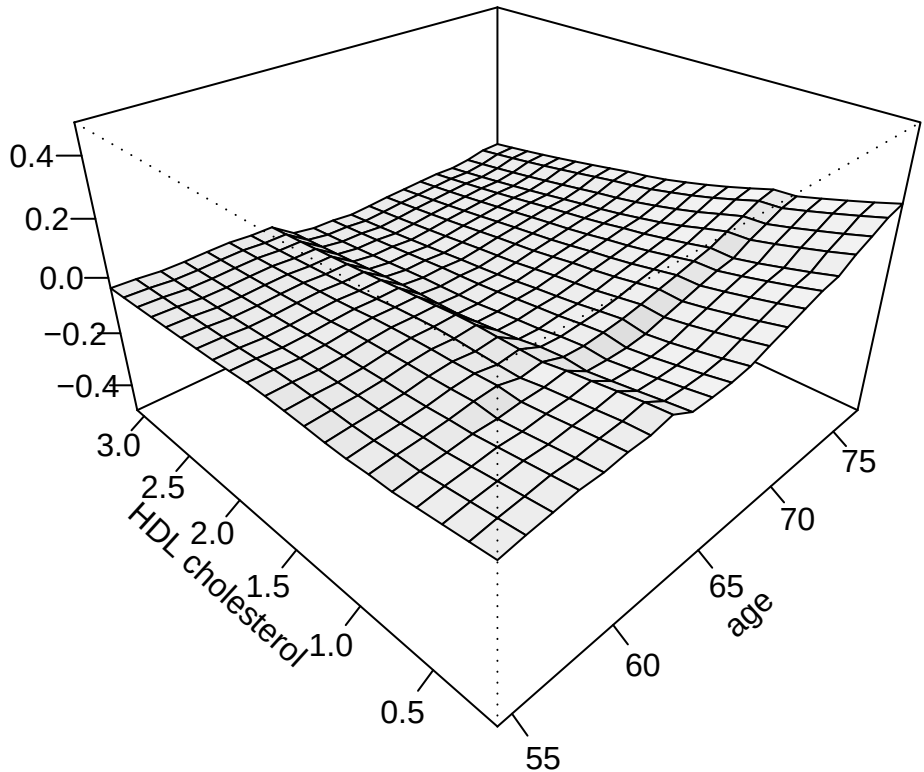
5. Unconstrained model for age and HDL: Posterior median 7 year risk



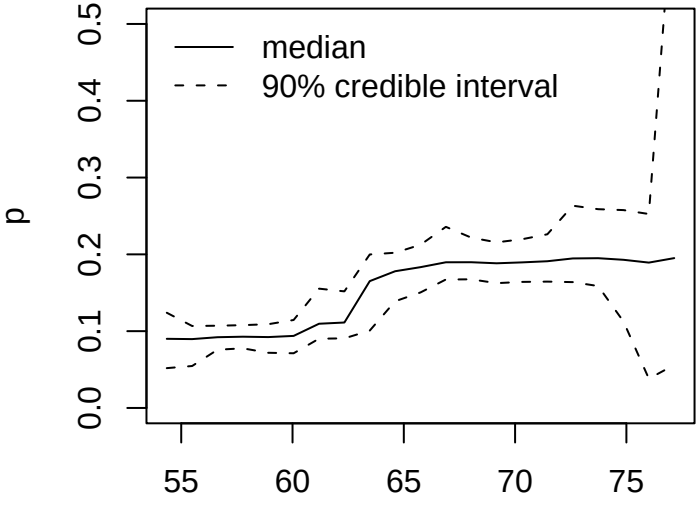
5. Unconstrained model for age and HDL: Length of 90% credible interval

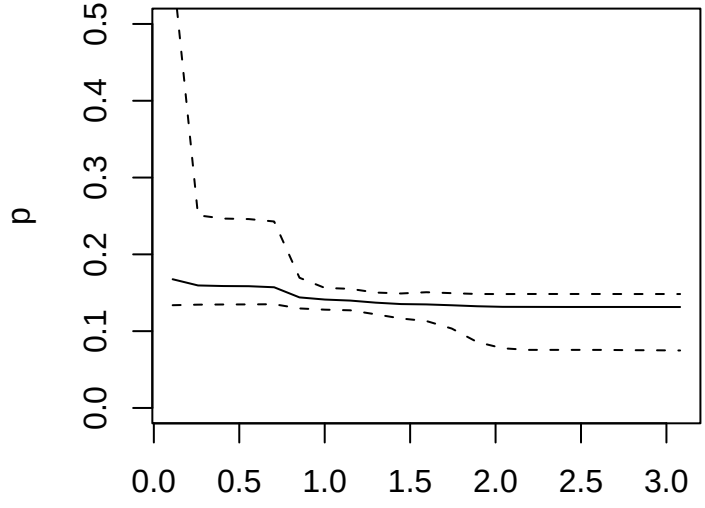


5. Unconstrained model for age and HDL: Difference to proportional hazards model

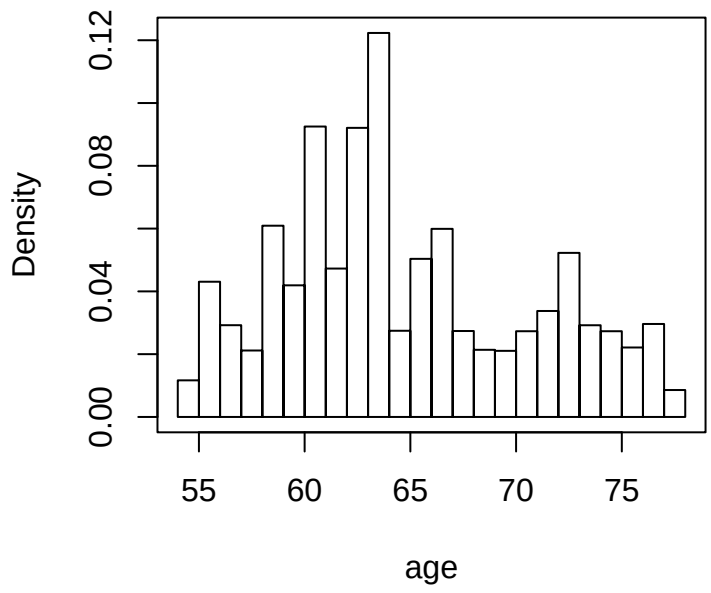


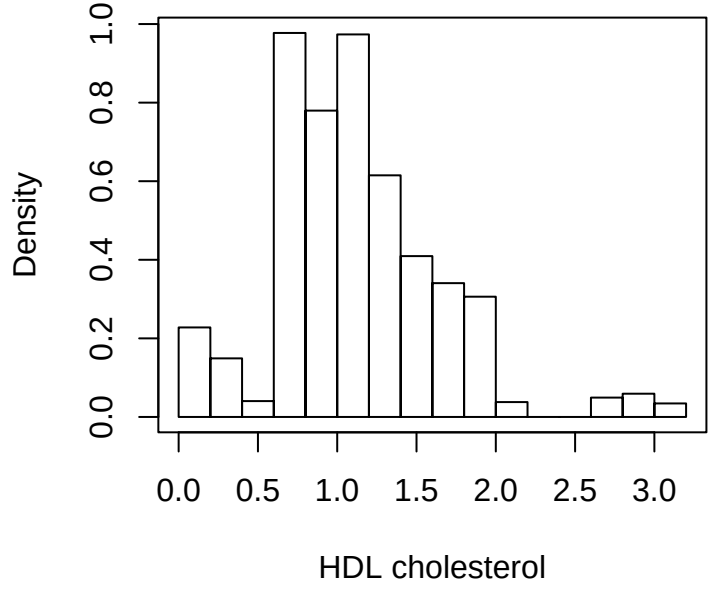
5. Unconstrained model for age and HDL: One−dimensional projections

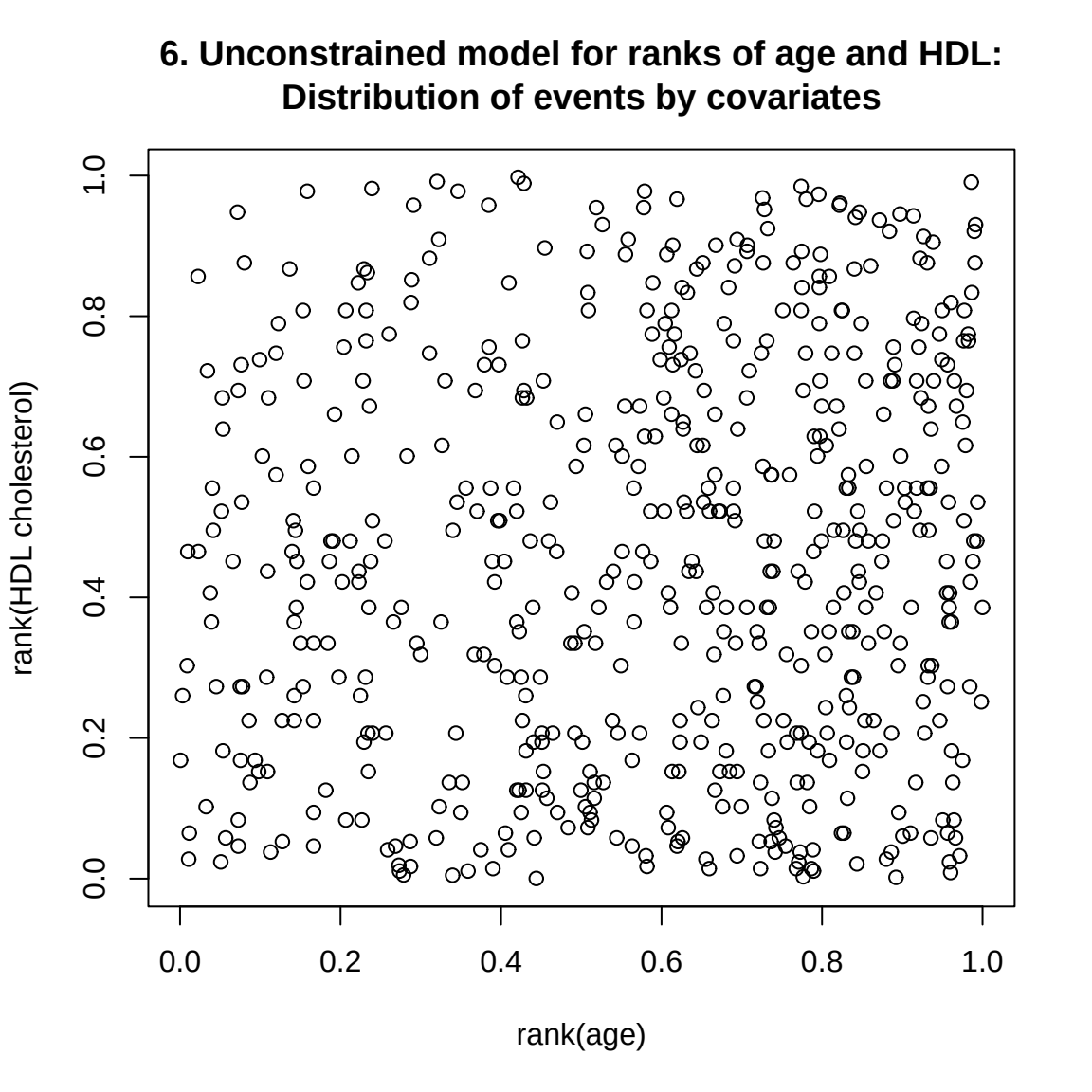


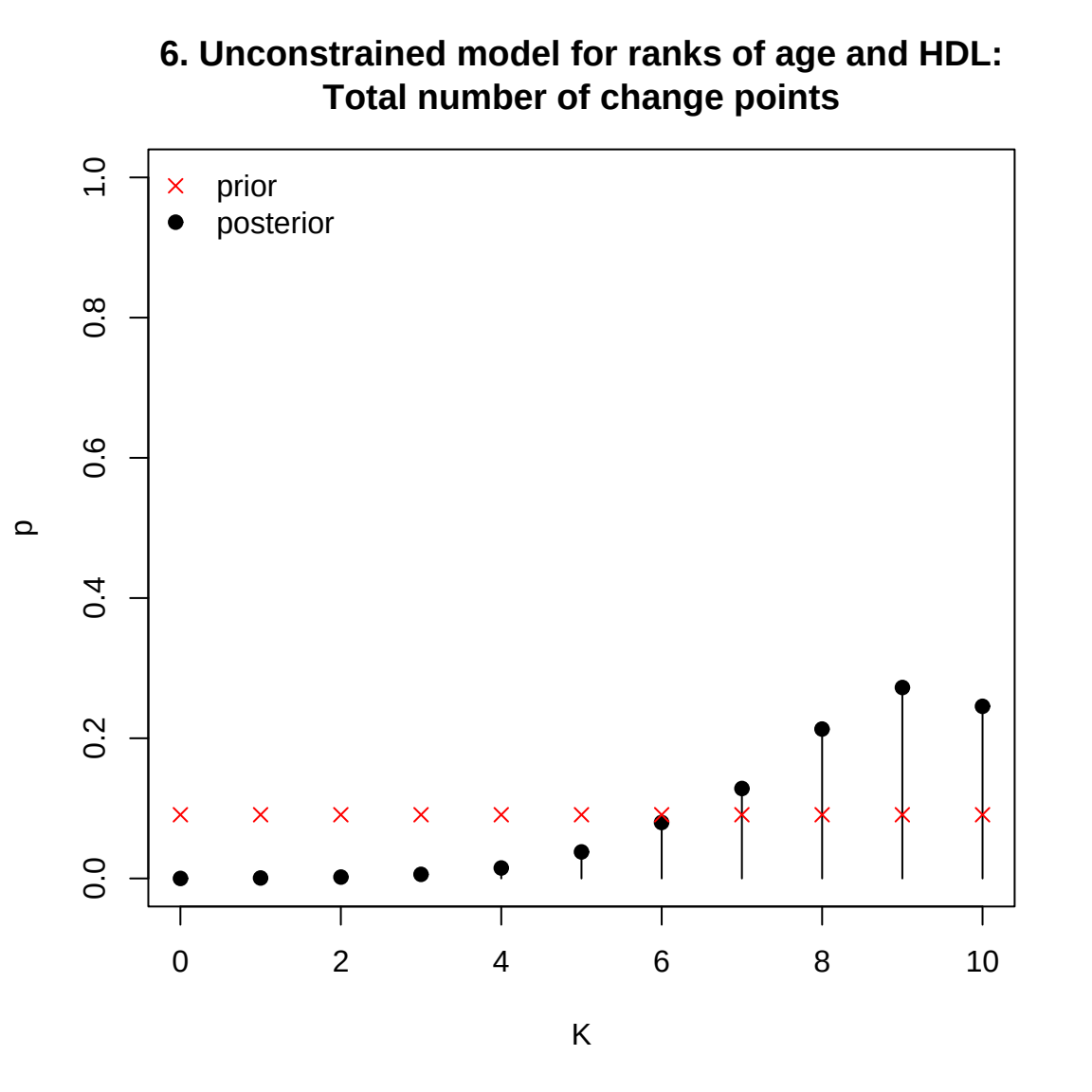


5. Unconstrained model for age and HDL: Distributions for change point positions

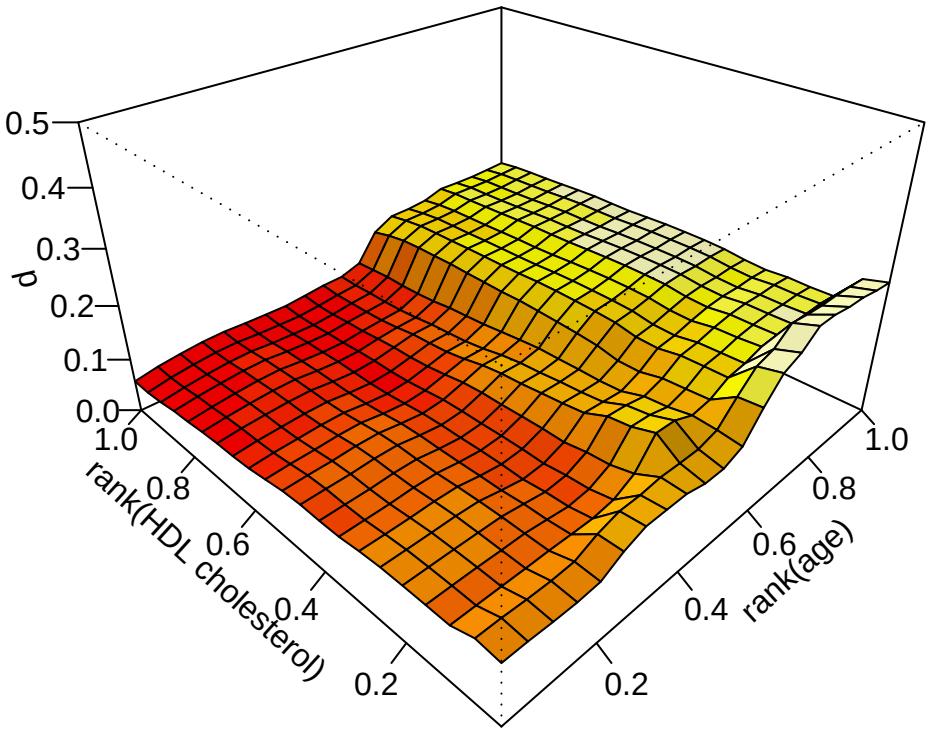




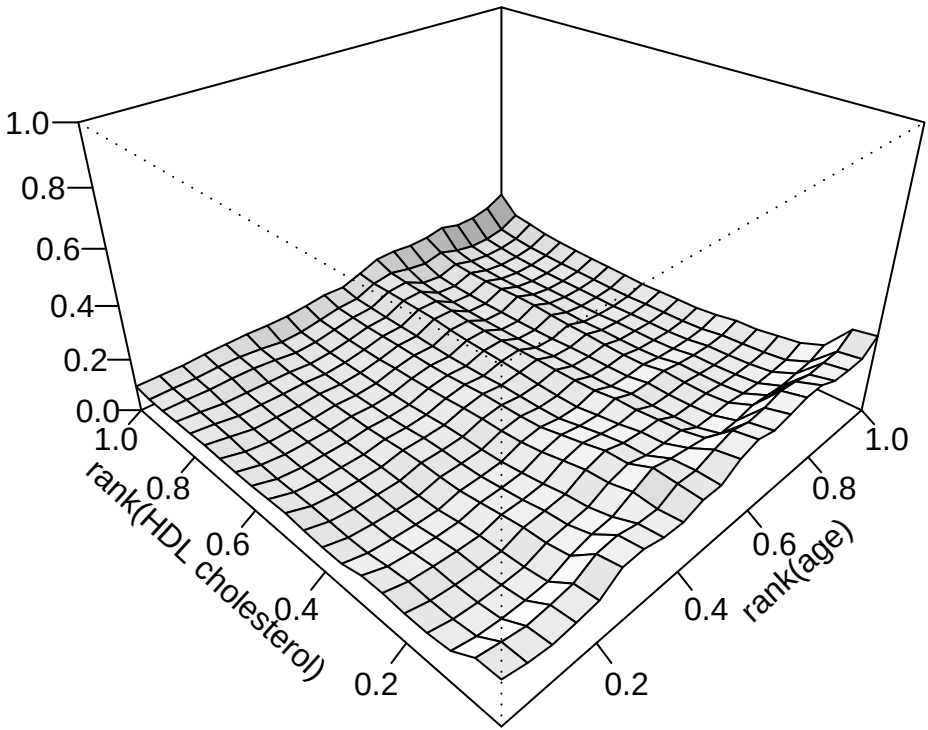

6. Unconstrained model for ranks of age and HDL:
Distribution of events by covariates
rank(HDL cholesterol)
0.0
0.2
0.4
0.6
0.8
1.0
rank(age)
0.0
0.2
0.4
0.6
0.8
1.0


6. Unconstrained model for ranks of age and HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

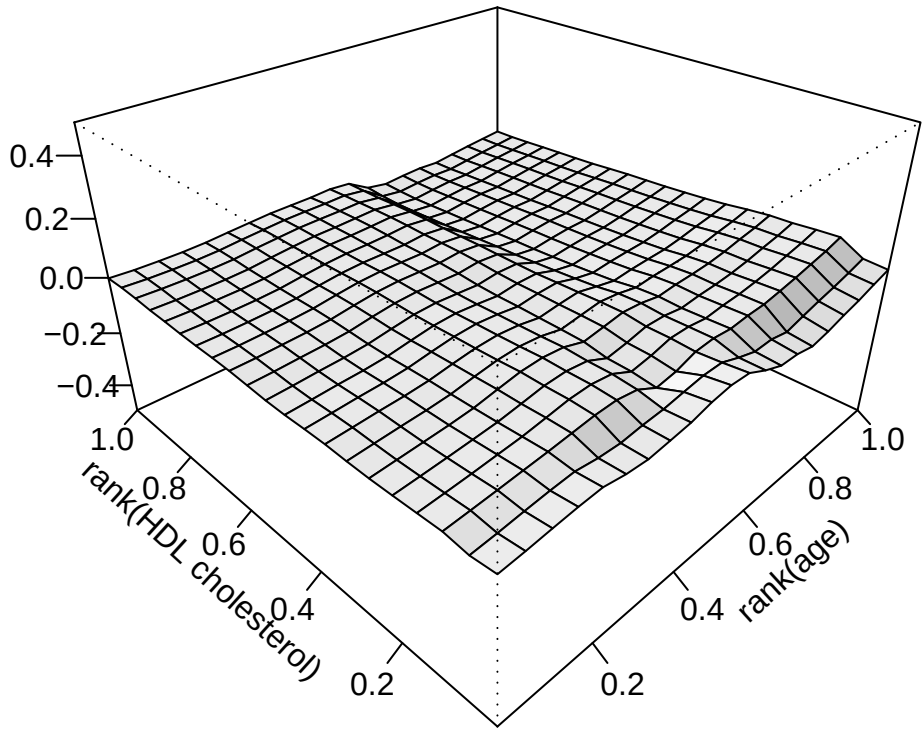
6. Unconstrained model for ranks of age and HDL: Posterior median 7 year risk



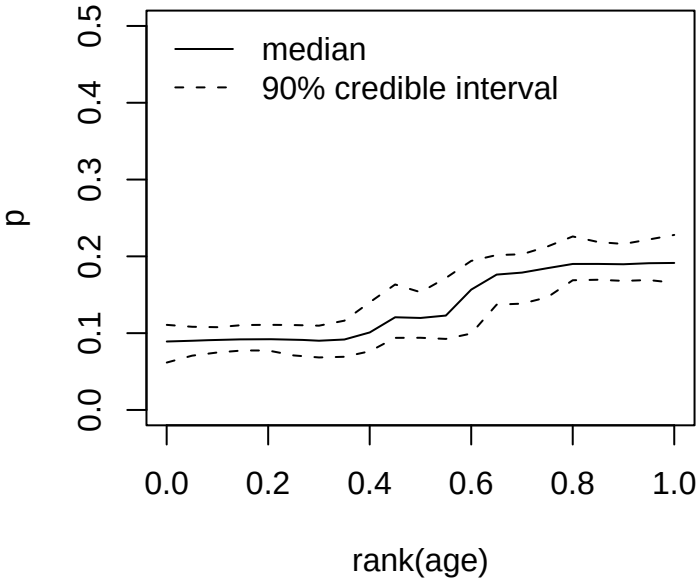
6. Unconstrained model for ranks of age and HDL: Length of 90% credible interval

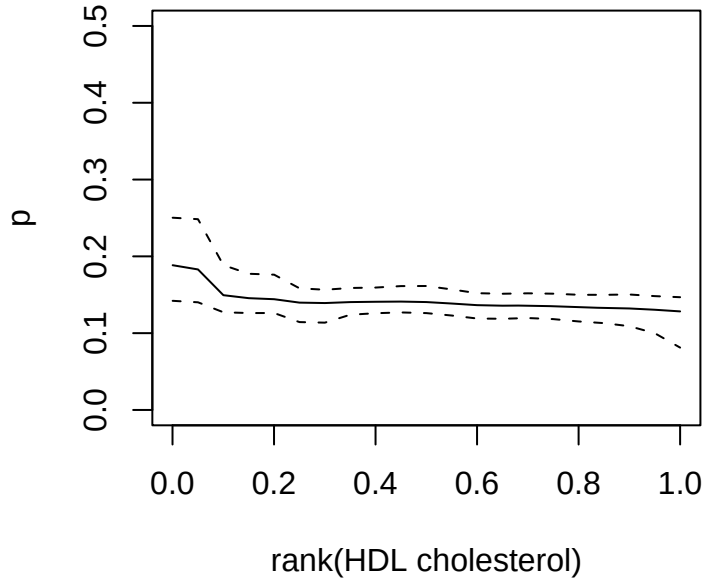


6. Unconstrained model for ranks of age and HDL: Difference to proportional hazards model

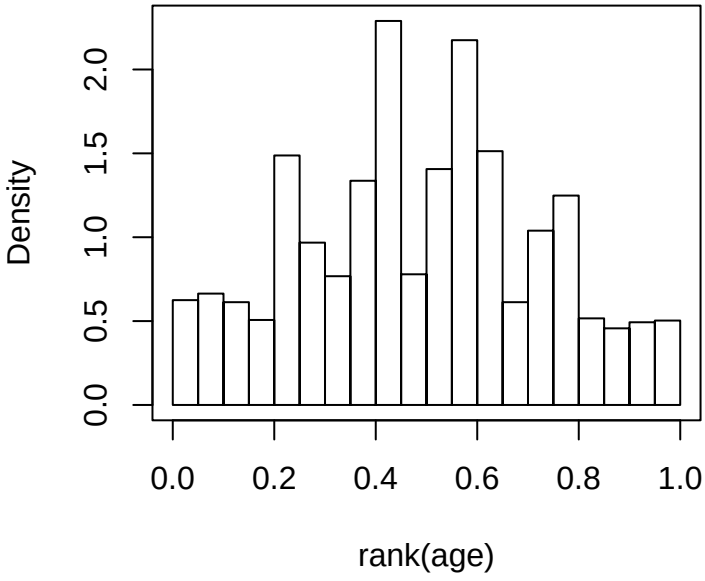


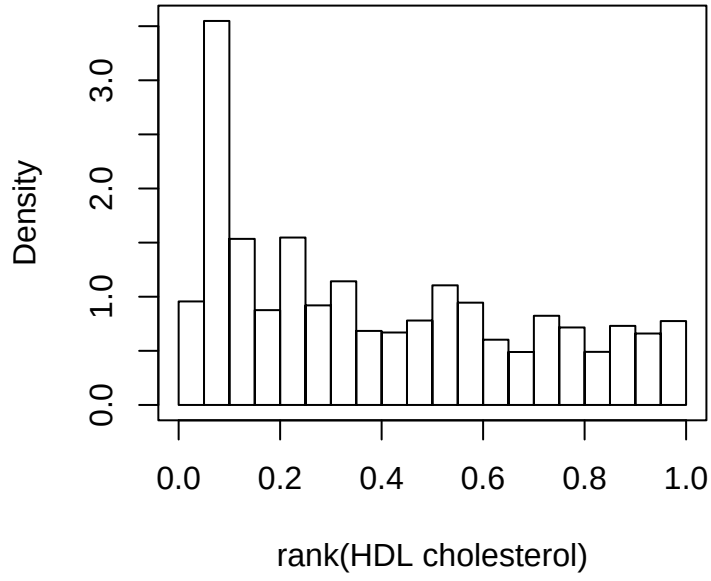
6. Unconstrained model for ranks of age and HDL: One−dimensional projections

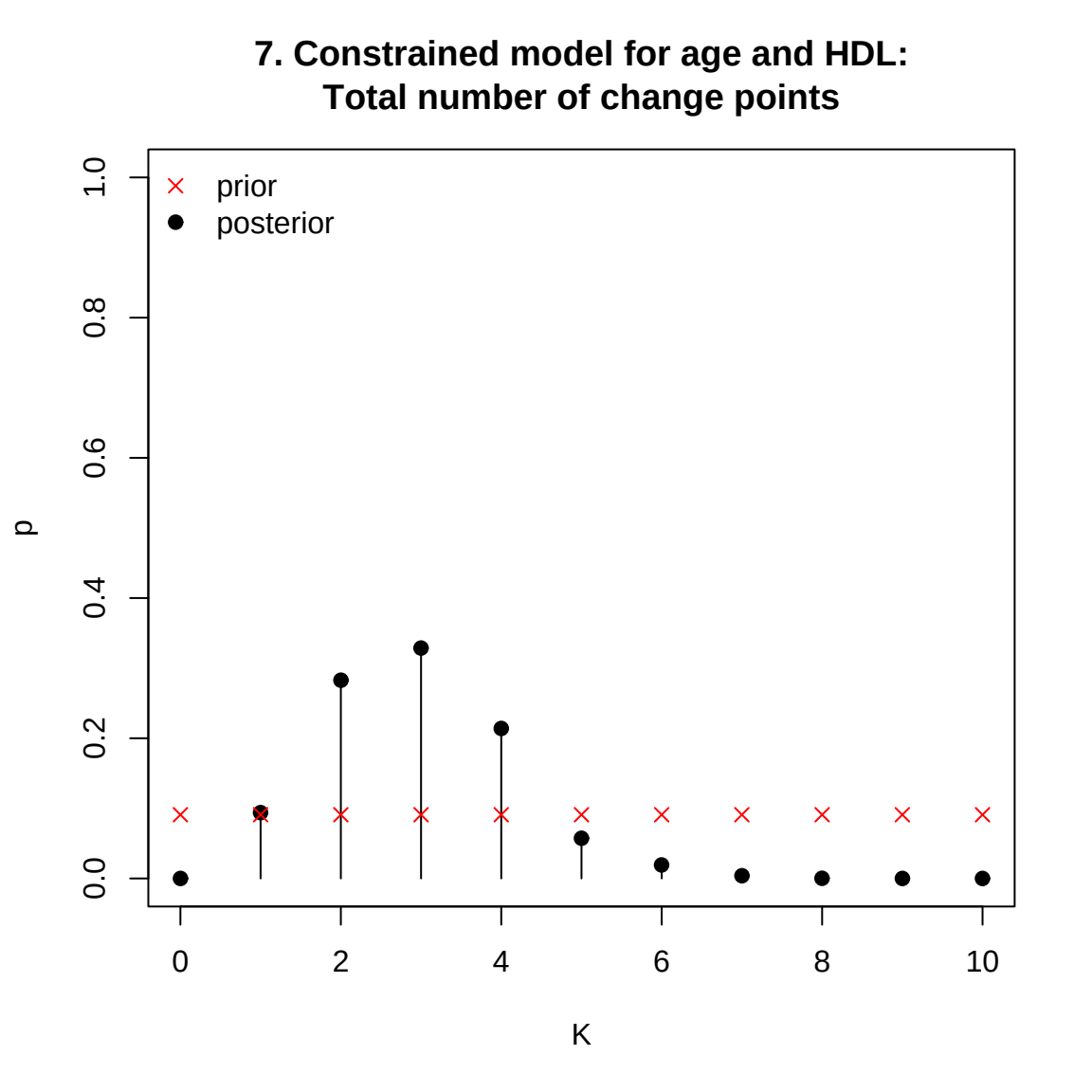




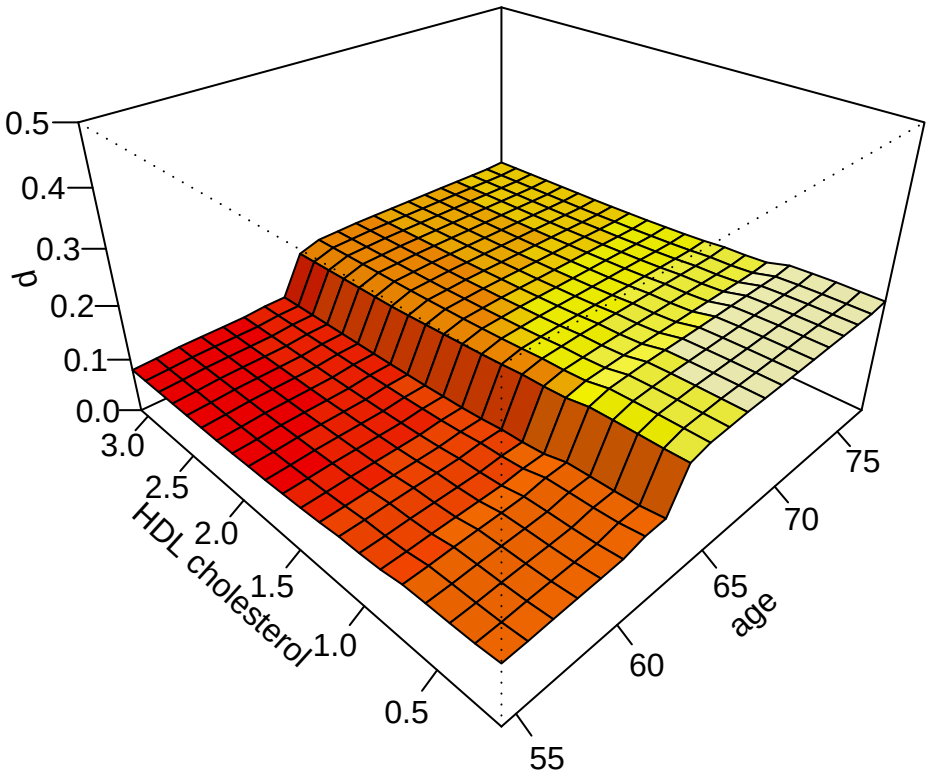
6. Unconstrained model for ranks of age and HDL: Distributions for change point positions



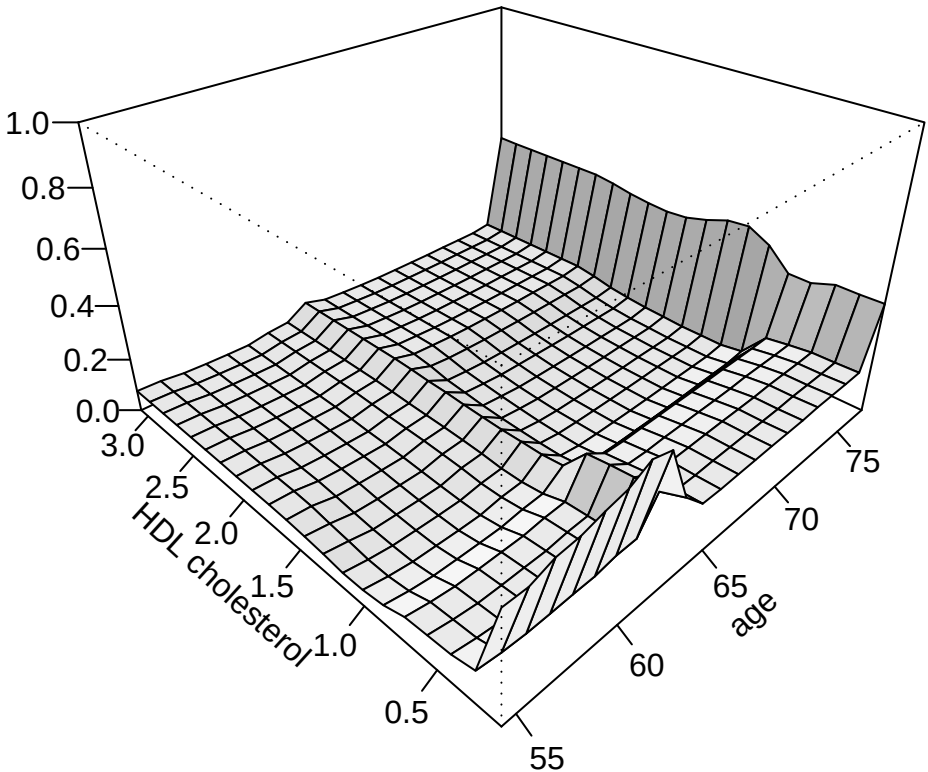



7. Constrained model for age and HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

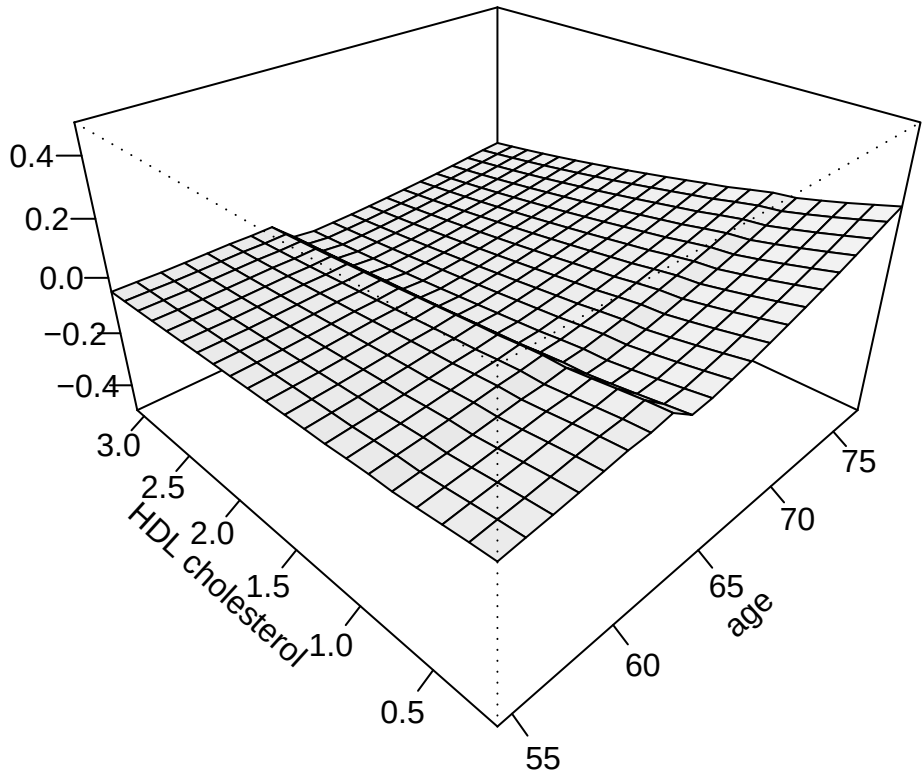
7. Constrained model for age and HDL: Posterior median 7 year risk



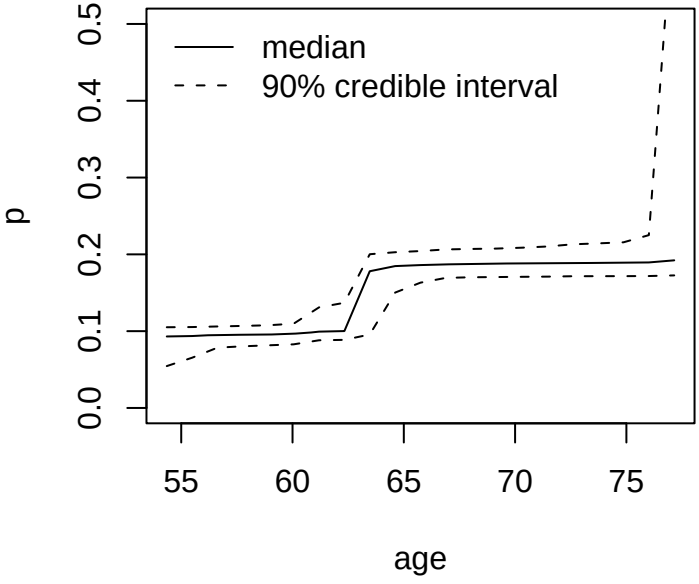
7. Constrained model for age and HDL: Length of 90% credible interval

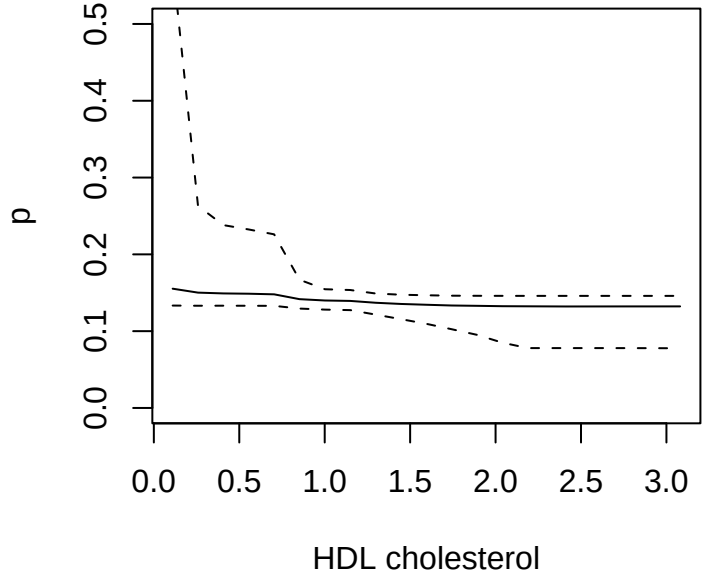


7. Constrained model for age and HDL: Difference to proportional hazards model

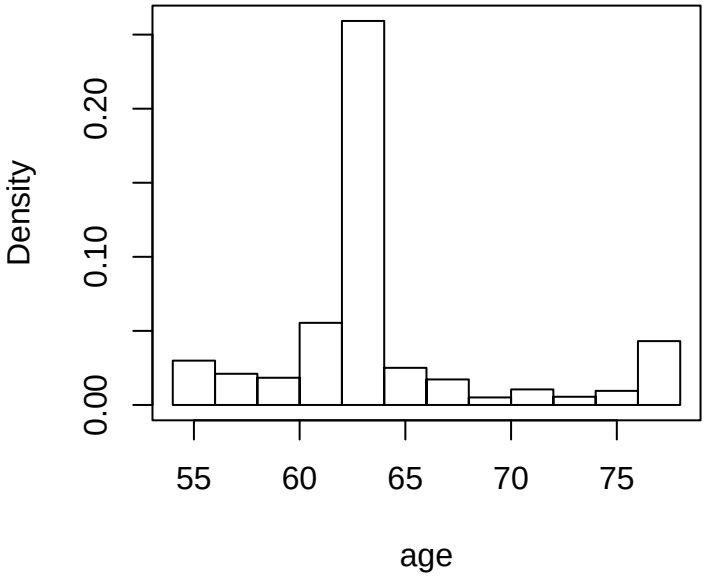


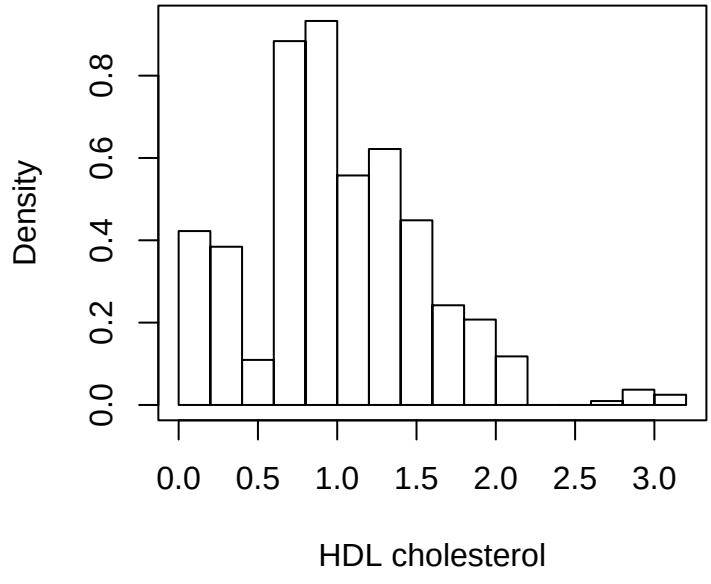
7. Constrained model for age and HDL: One−dimensional projections

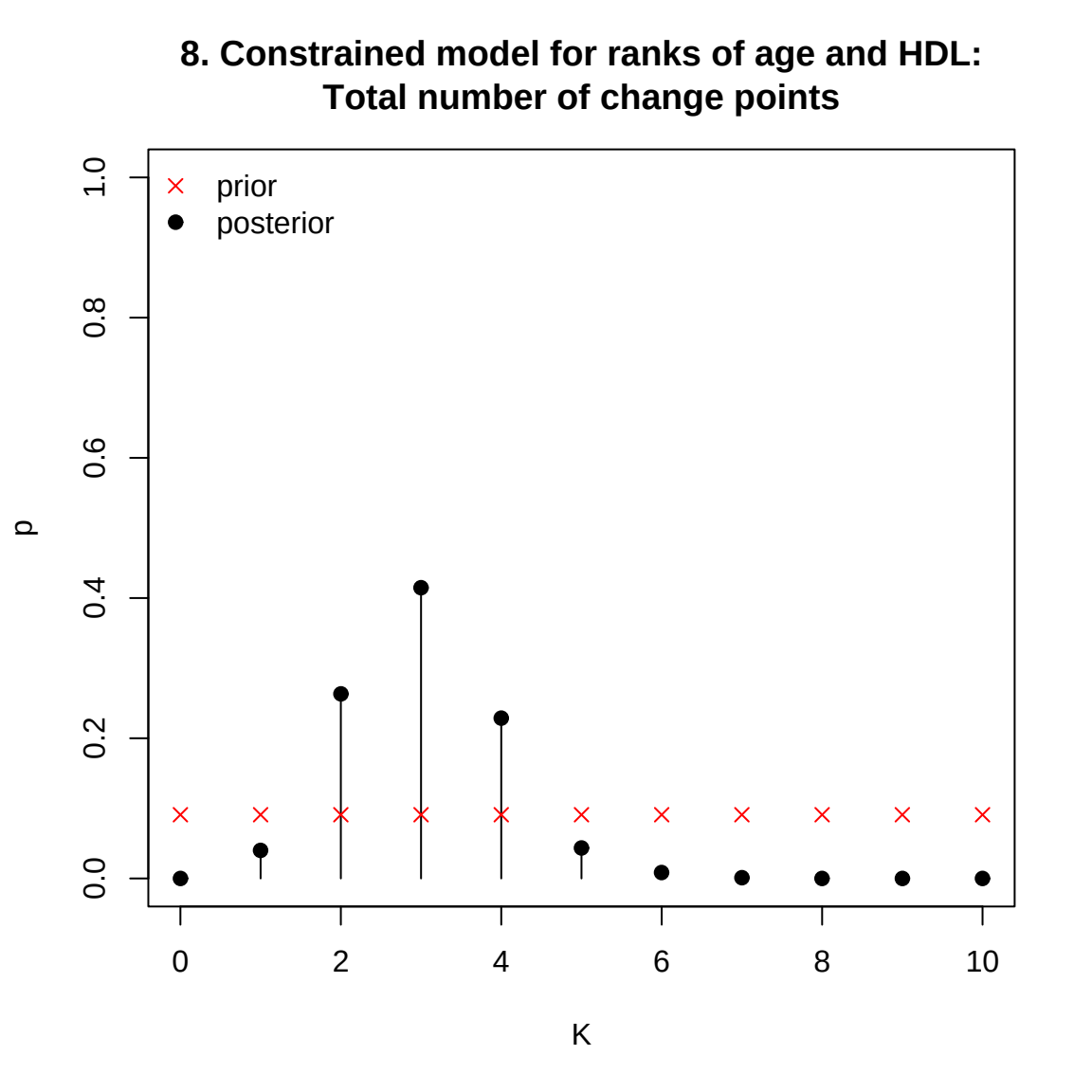




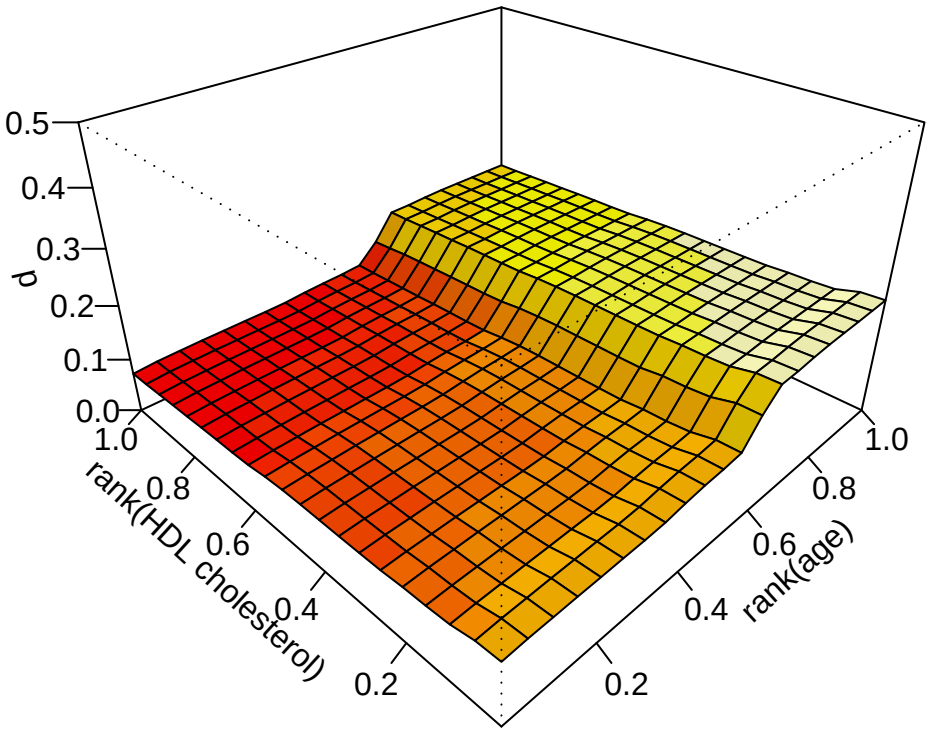
7. Constrained model for age and HDL: Distributions for change point positions



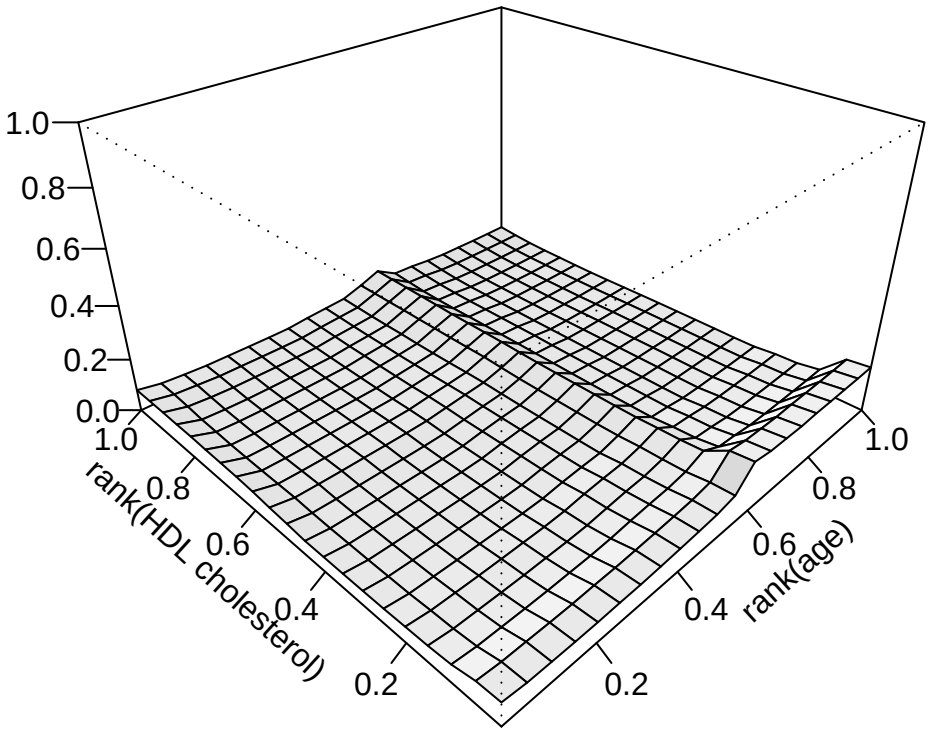



8. Constrained model for ranks of age and HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

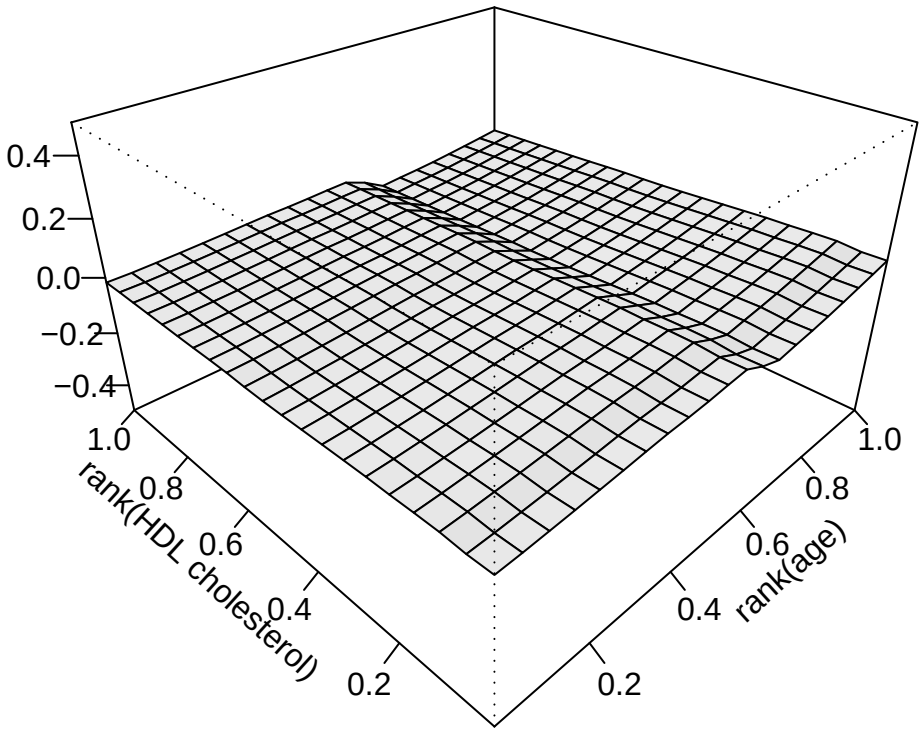
8. Constrained model for ranks of age and HDL: Posterior median 7 year risk



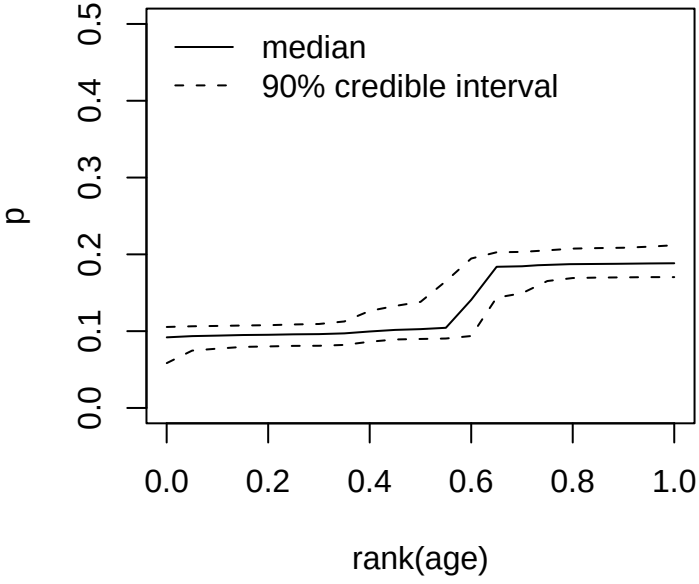
8. Constrained model for ranks of age and HDL: Length of 90% credible interval

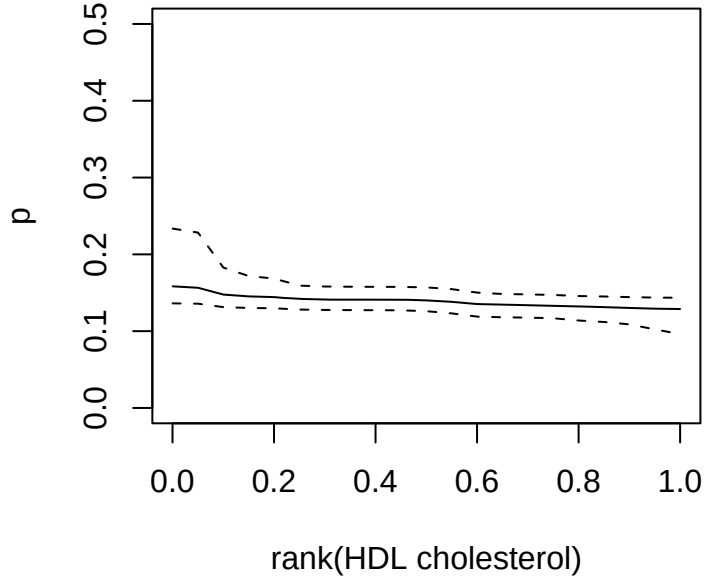


8. Constrained model for ranks of age and HDL: Difference to proportional hazards model

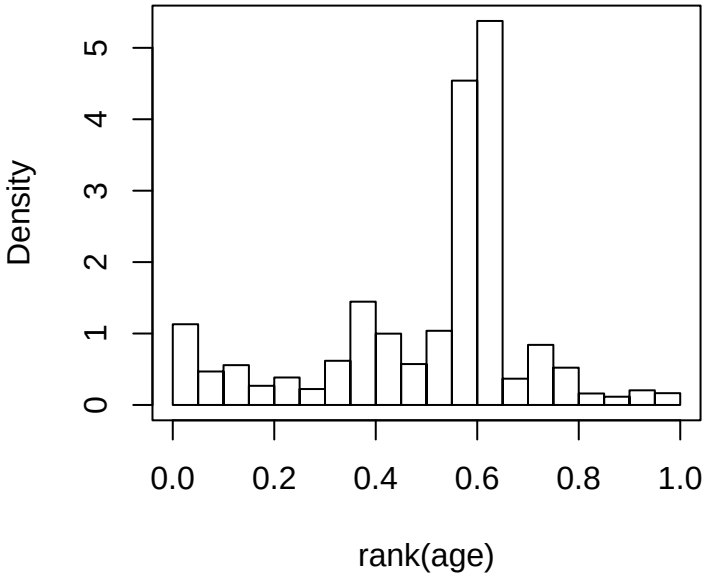


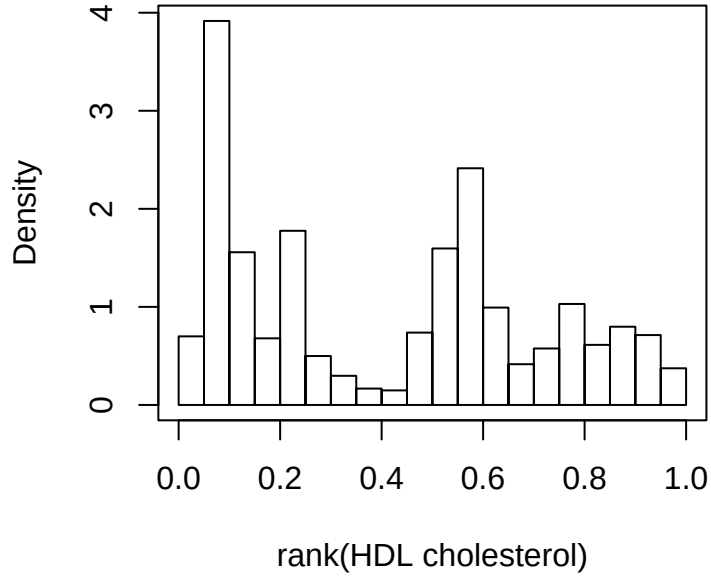
8. Constrained model for ranks of age and HDL: One−dimensional projections



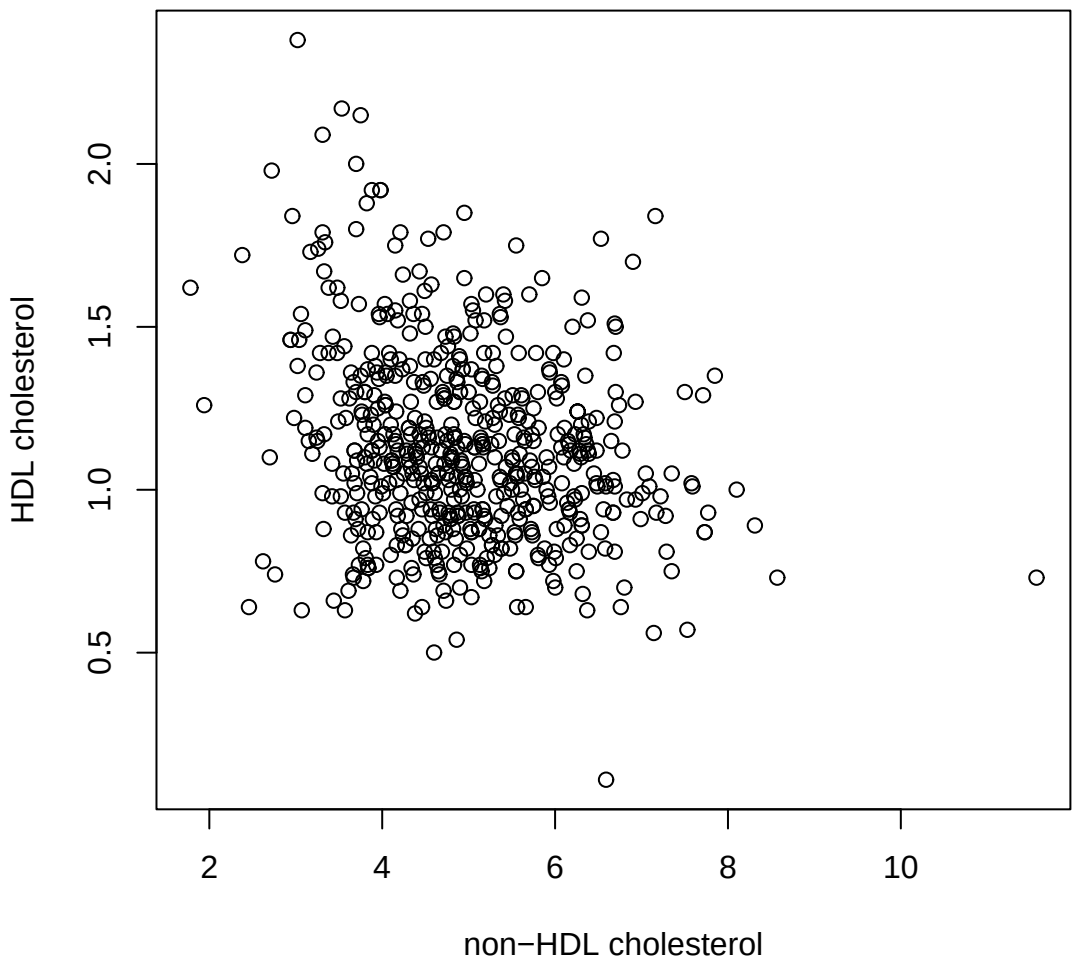


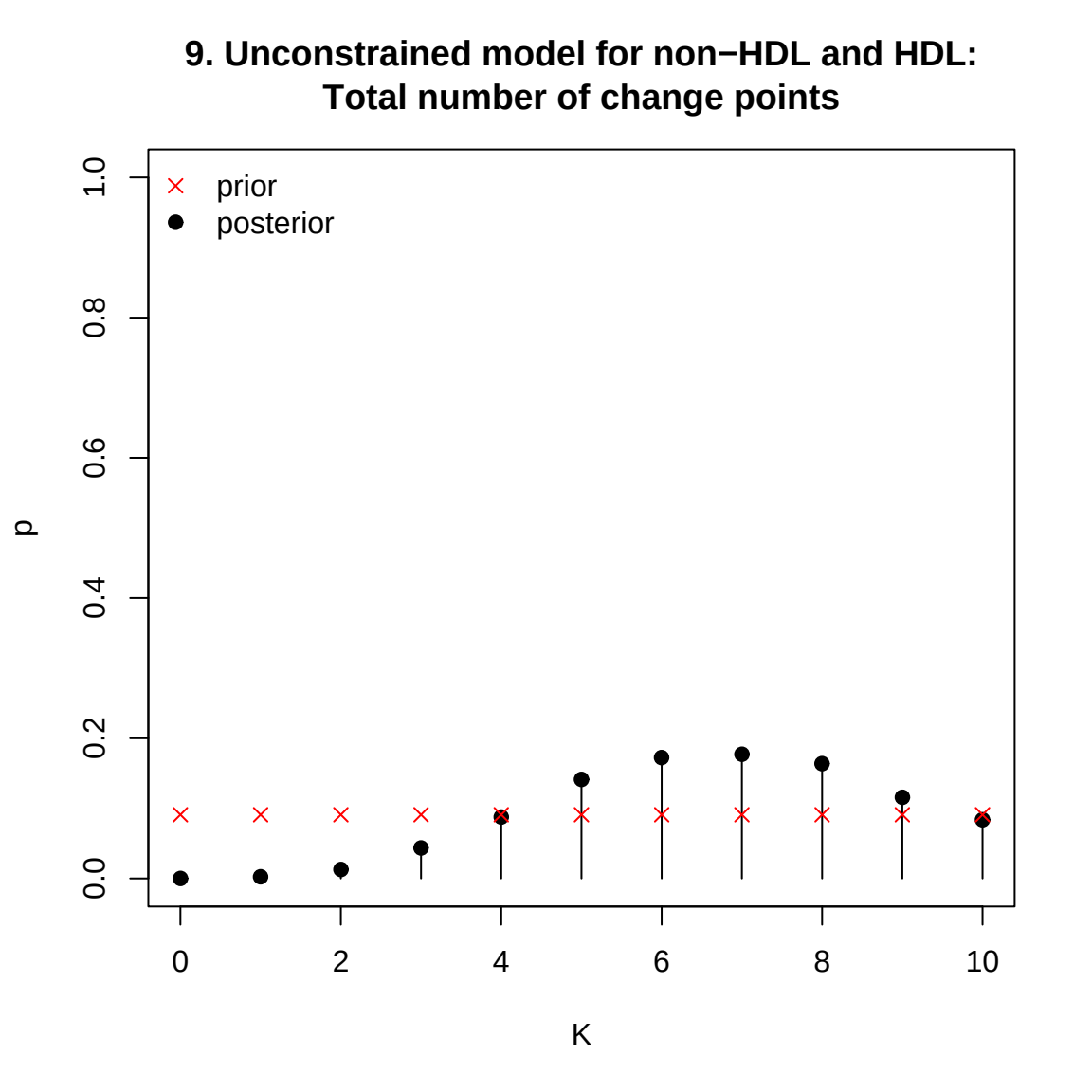
8. Constrained model for ranks of age and HDL: Distributions for change point positions



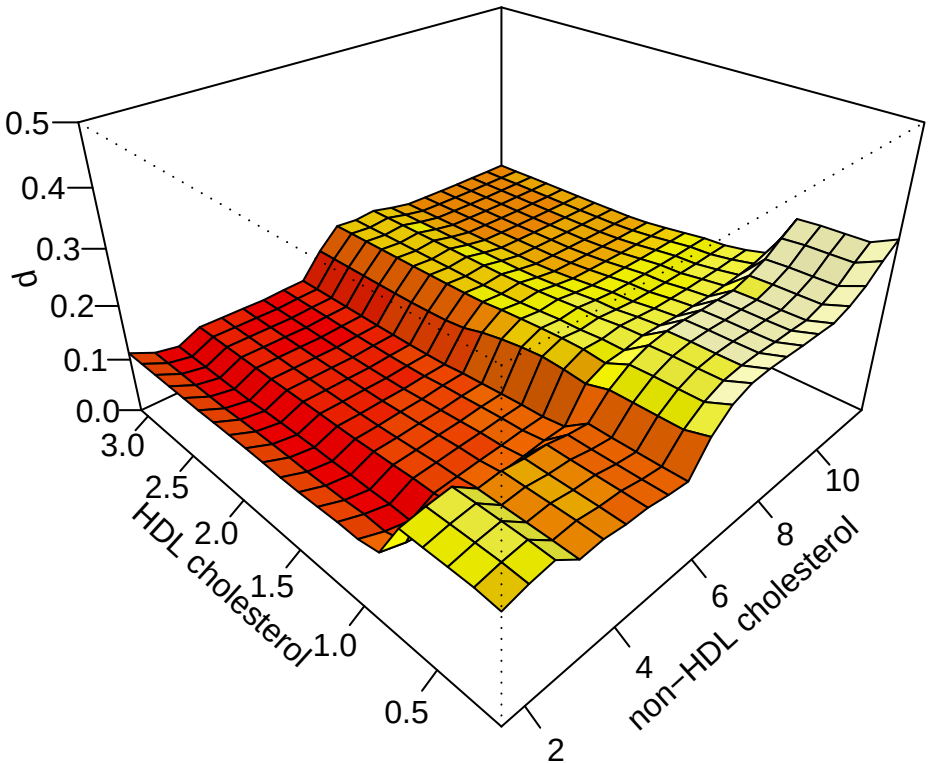


9. Unconstrained model for non−HDL and HDL: Distribution of events by covariates

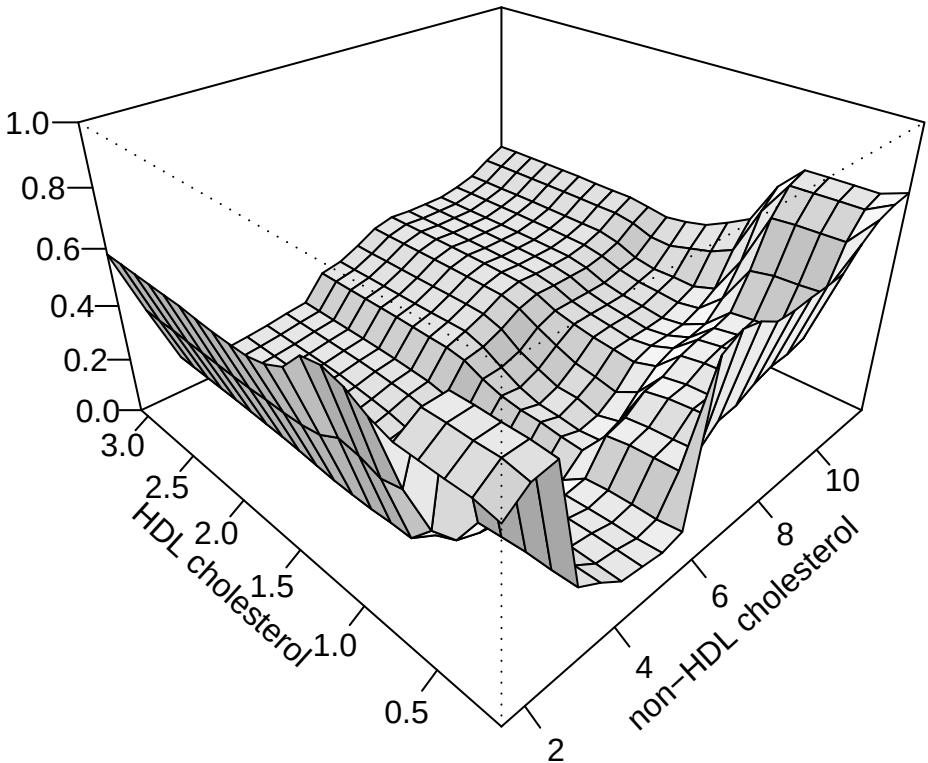



9. Unconstrained model for non−HDL and HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

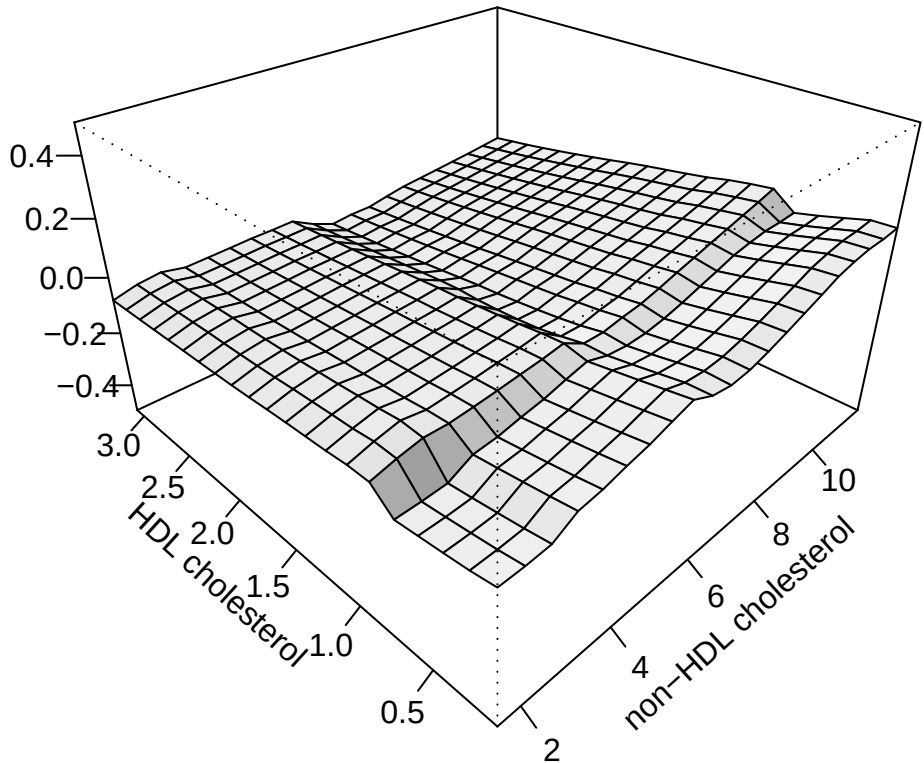
9. Unconstrained model for non−HDL and HDL: Posterior median 7 year risk



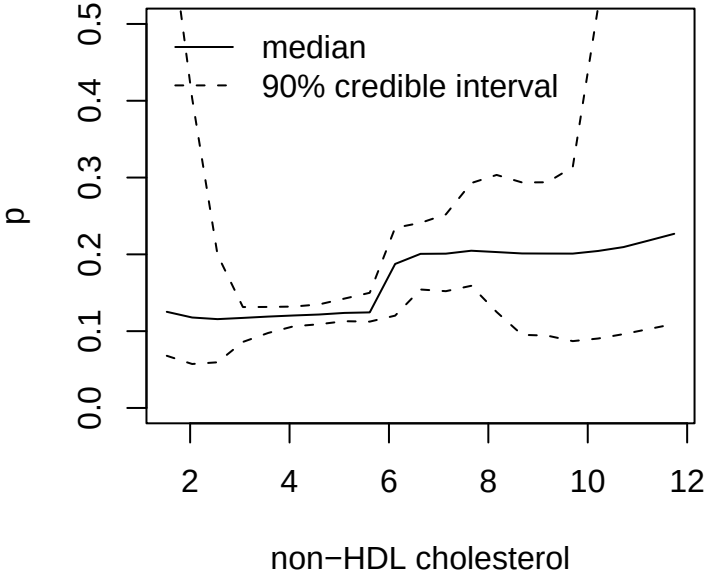
9. Unconstrained model for non−HDL and HDL: Length of 90% credible interval

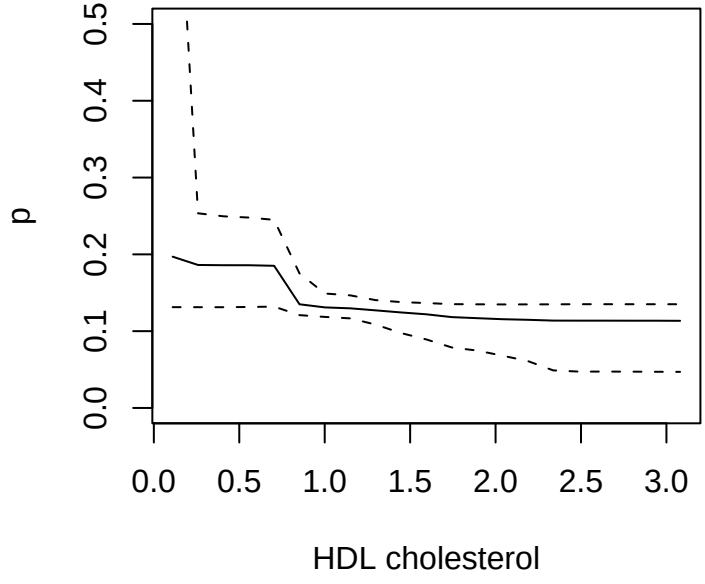


9. Unconstrained model for non−HDL and HDL: Difference to proportional hazards model

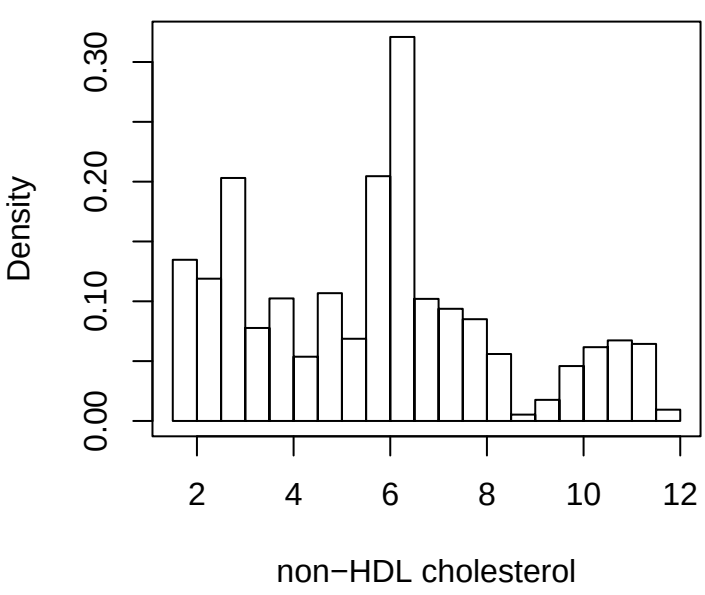


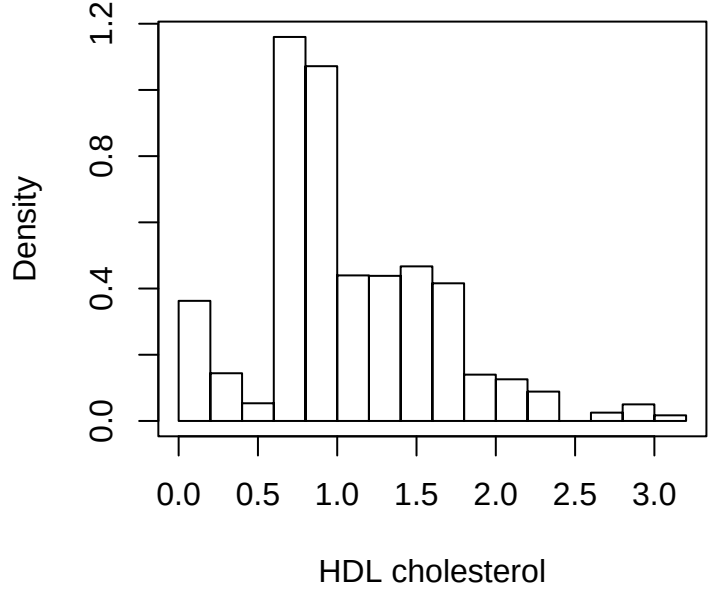
9. Unconstrained model for non−HDL and HDL: One−dimensional projections



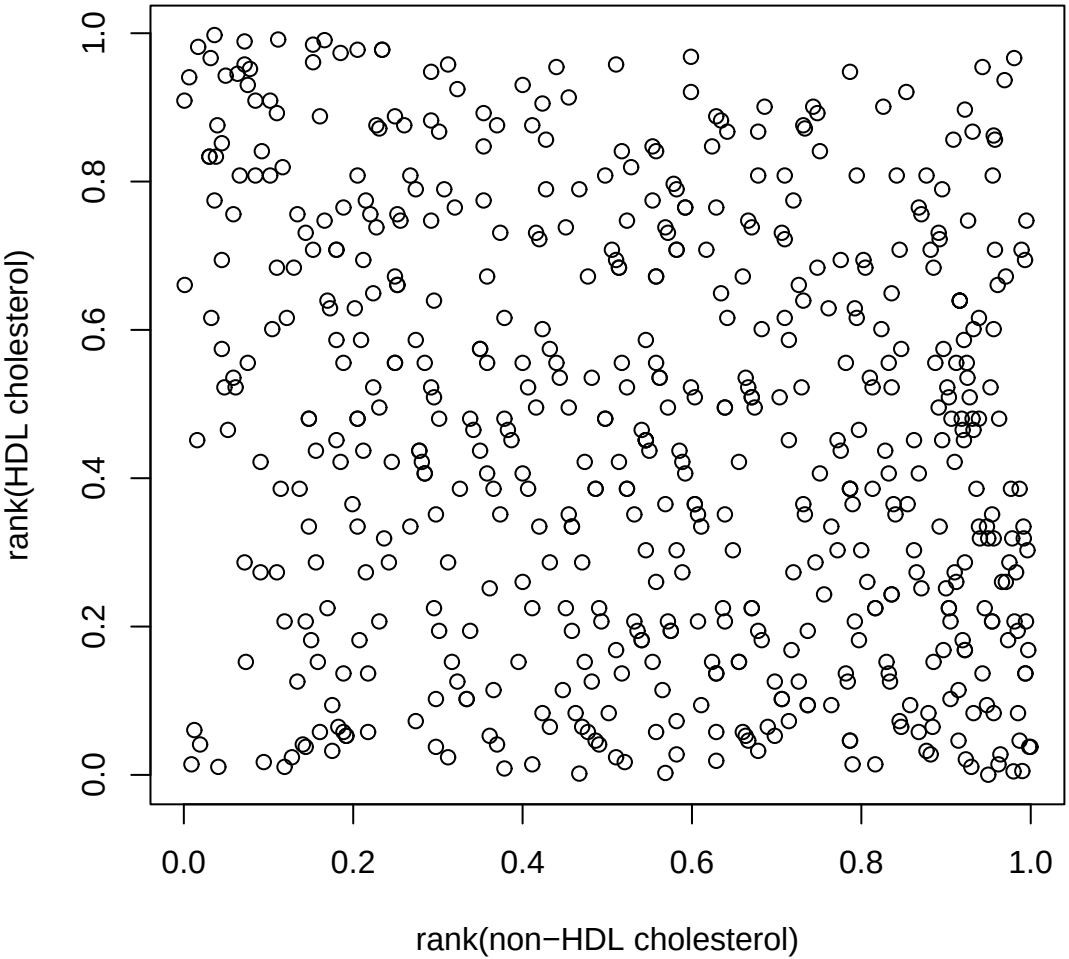


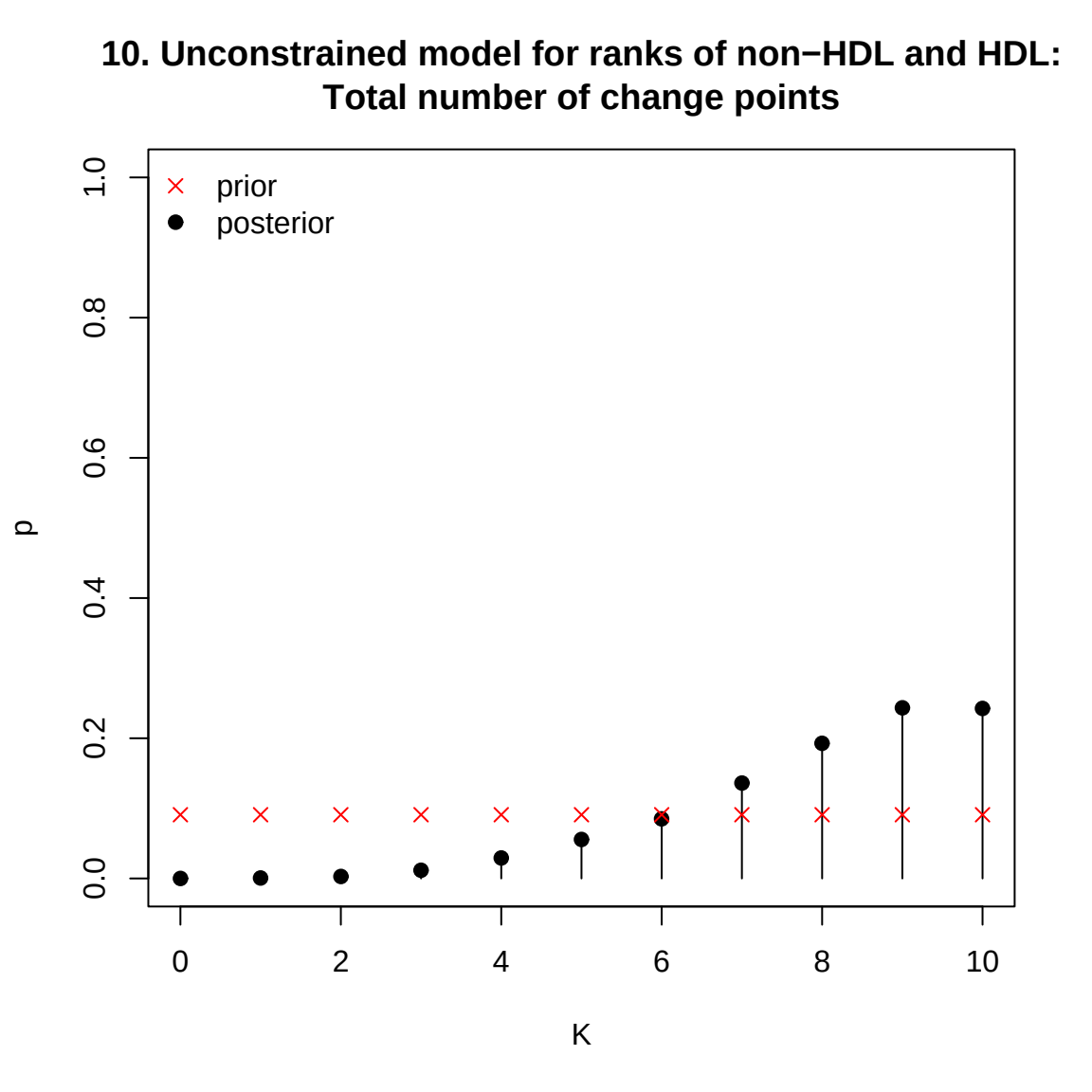
9. Unconstrained model for non−HDL and HDL: Distributions for change point positions



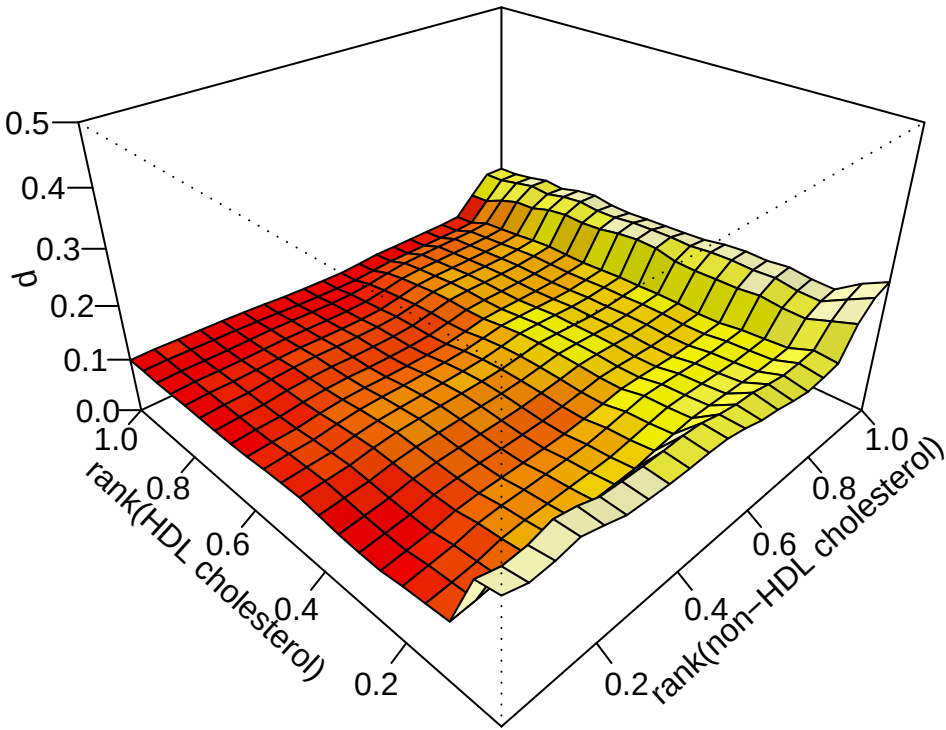


10. Unconstrained model for ranks of non−HDL and HDL: Distribution of events by covariates

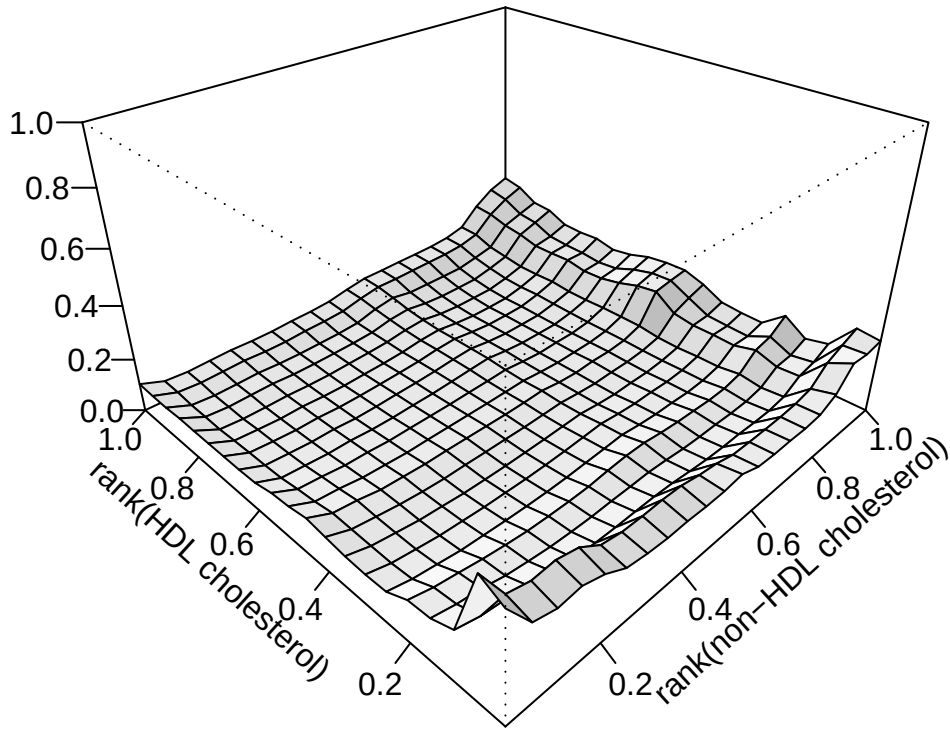



10. Unconstrained model for ranks of non−HDL and HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

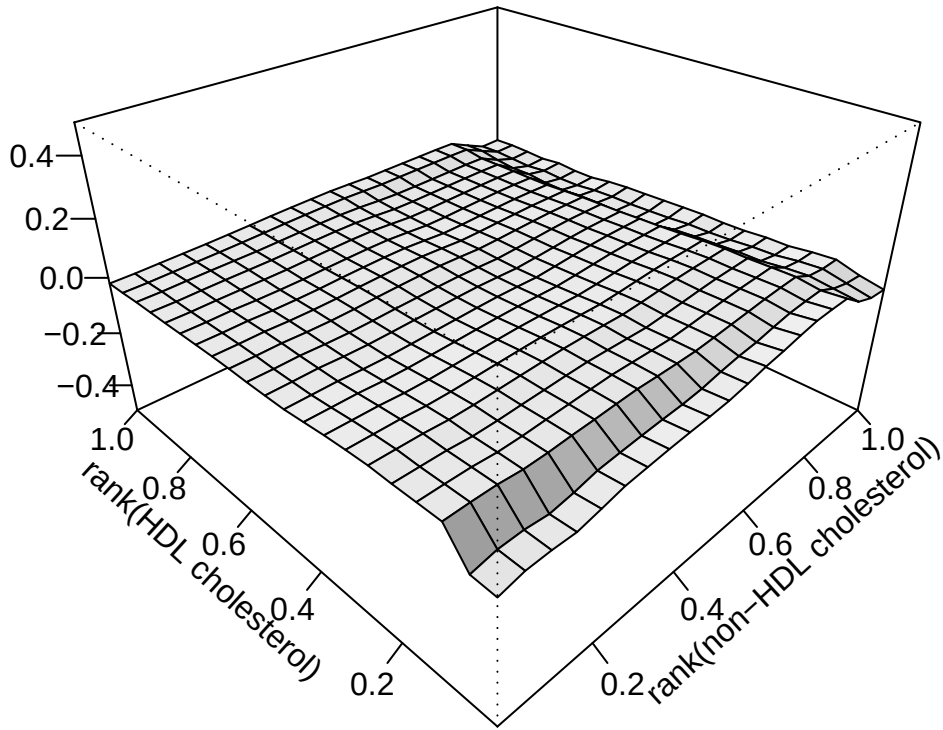
10. Unconstrained model for ranks of non−HDL and HDL: Posterior median 7 year risk



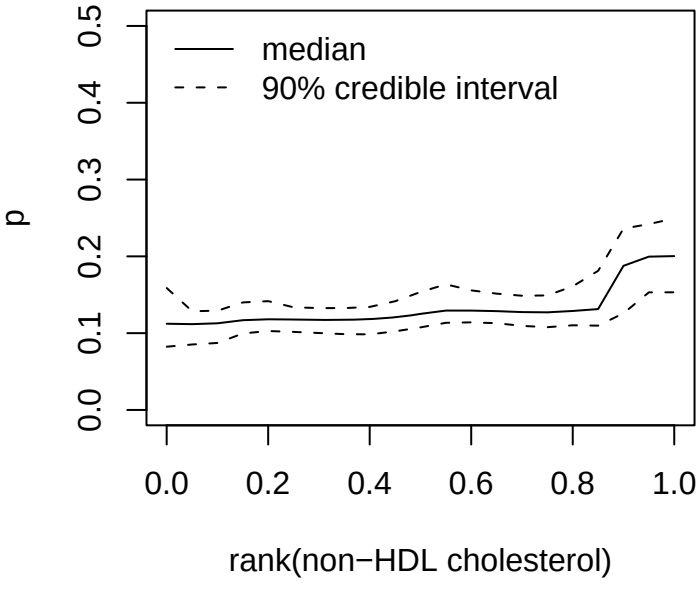
10. Unconstrained model for ranks of non−HDL and HDL: Length of 90% credible interval

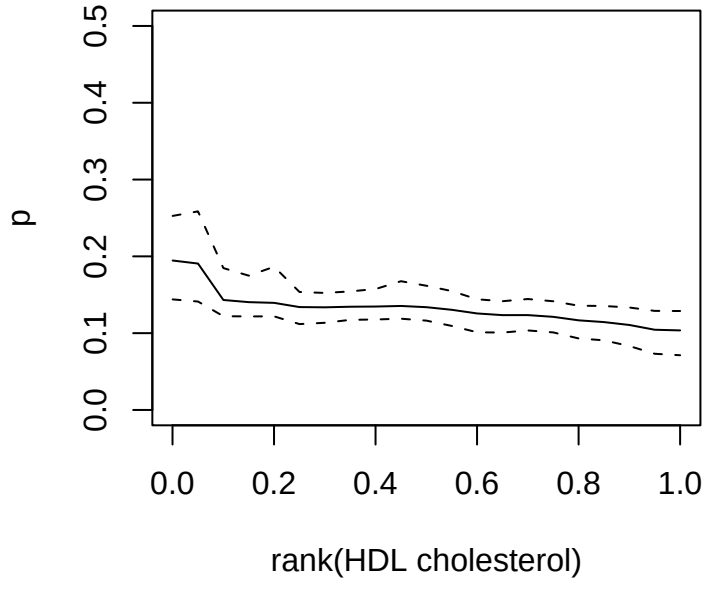


10. Unconstrained model for ranks of non−HDL and HDL: Difference to proportional hazards model

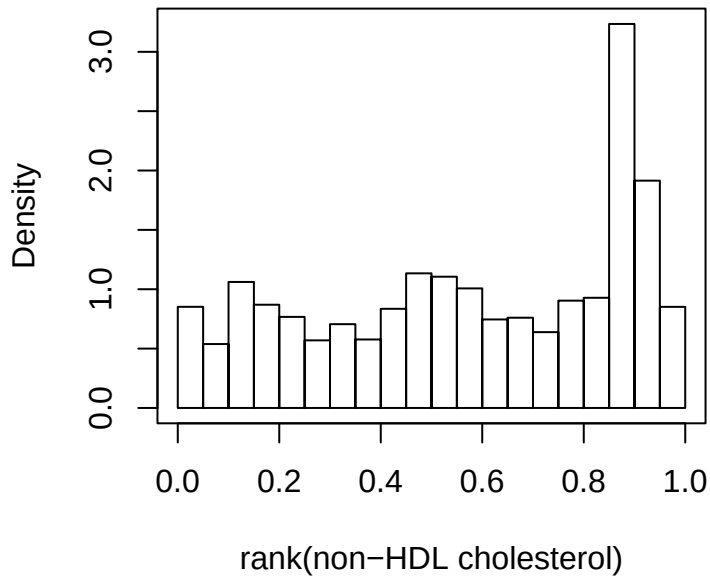


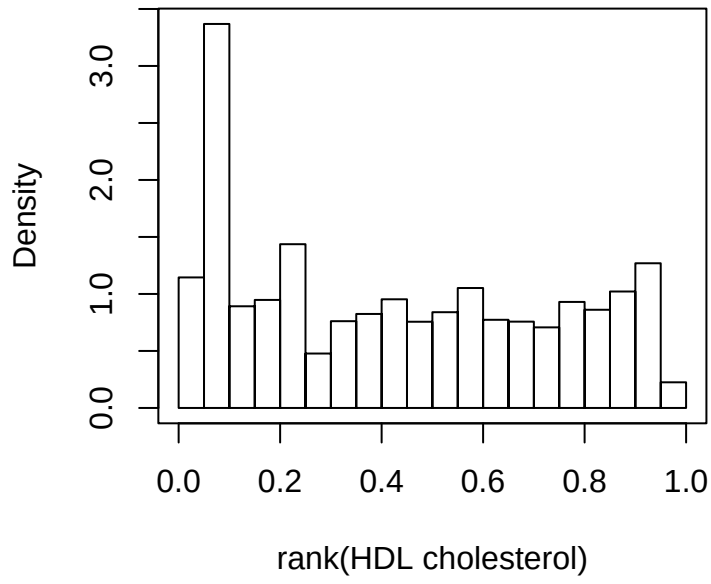
10. Unconstrained model for ranks of non−HDL and HDL: One−dimensional projections

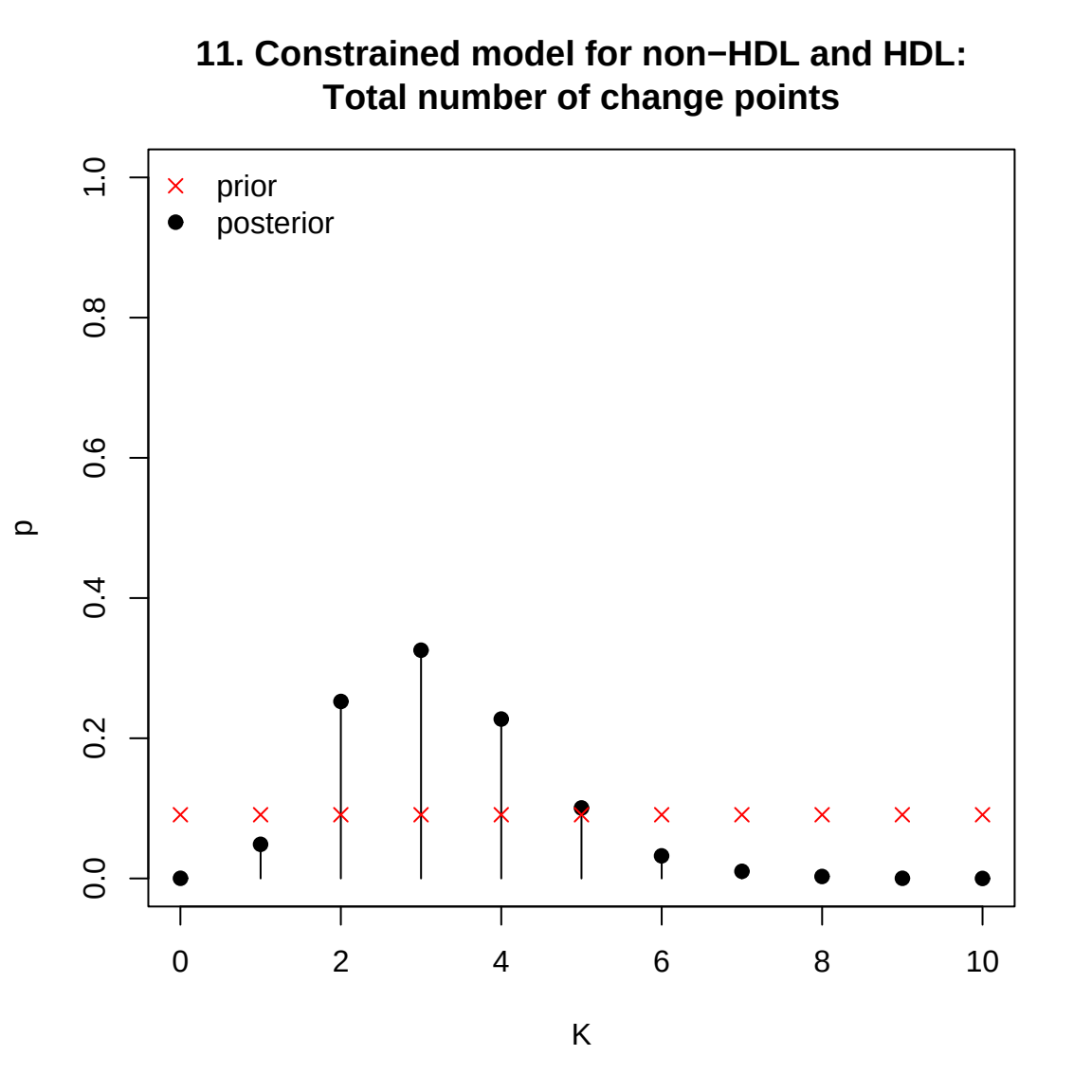




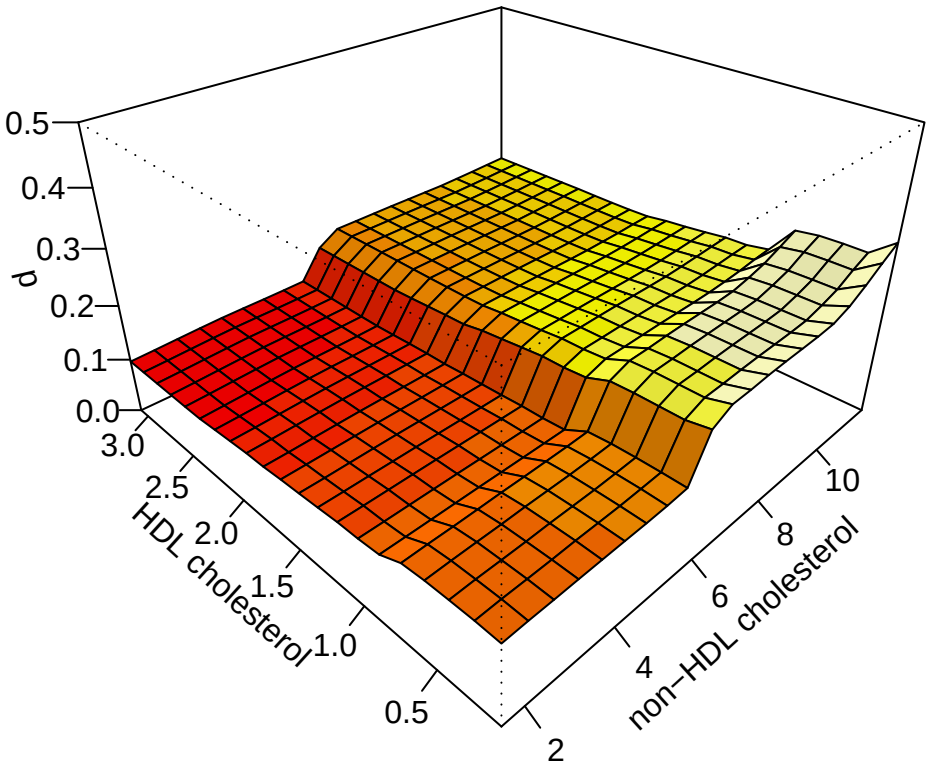
10. Unconstrained model for ranks of non−HDL and HDL: Distributions for change point positions



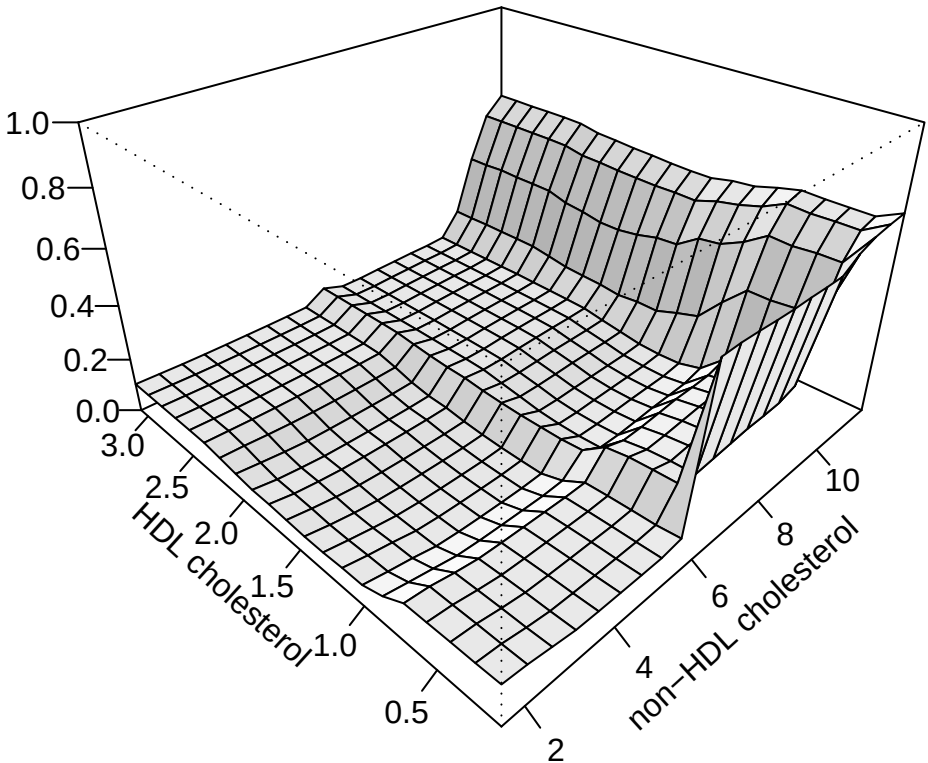



11. Constrained model for non−HDL and HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

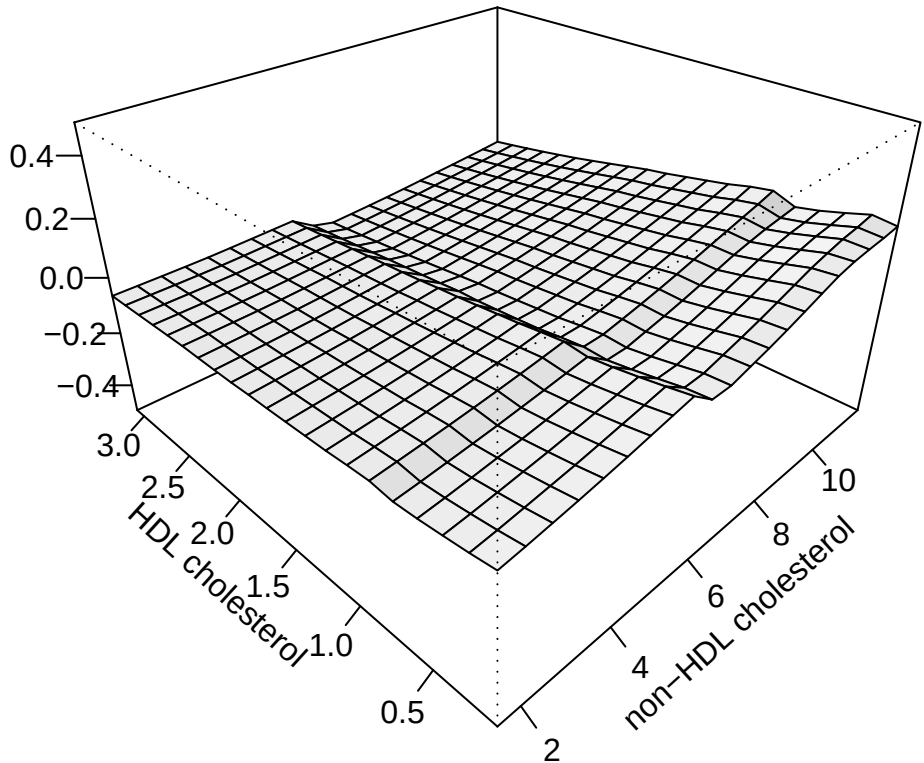
11. Constrained model for non−HDL and HDL: Posterior median 7 year risk



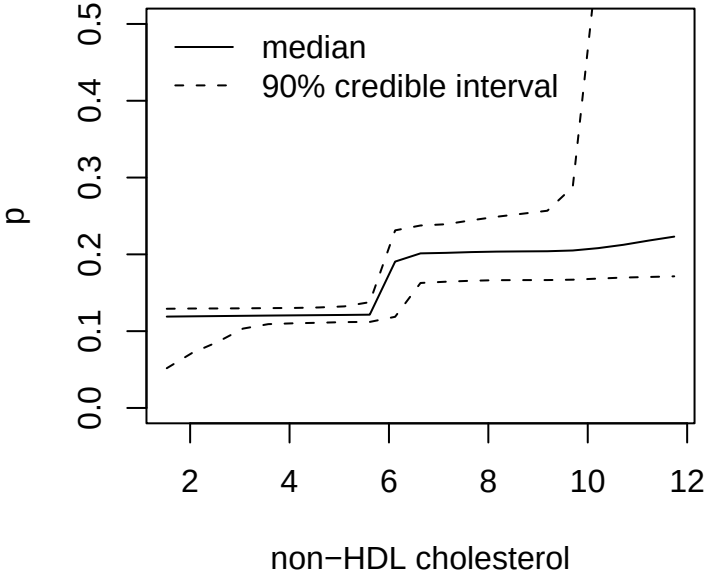
11. Constrained model for non−HDL and HDL: Length of 90% credible interval

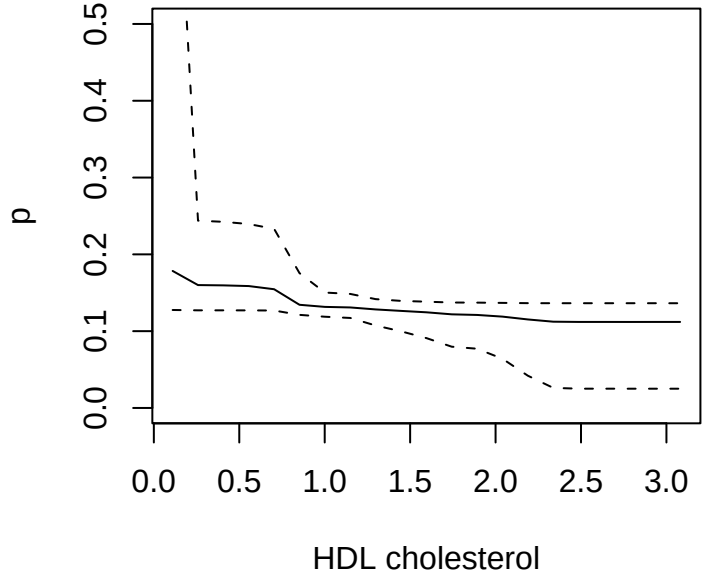


11. Constrained model for non−HDL and HDL: Difference to proportional hazards model

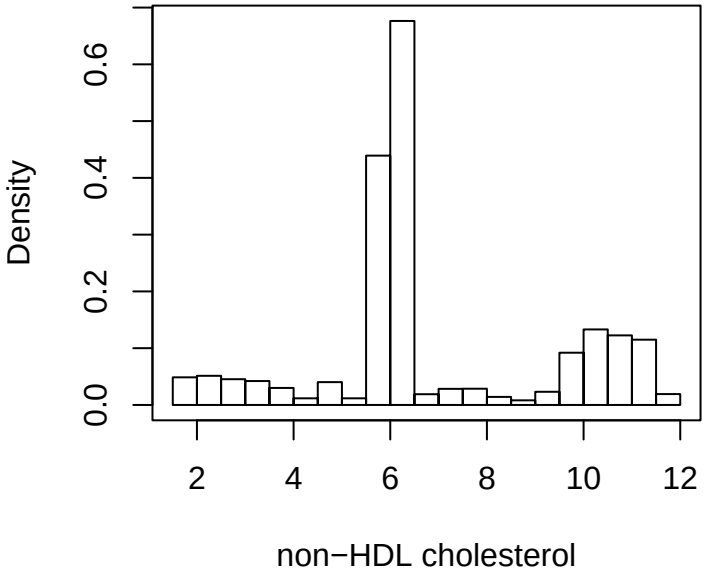


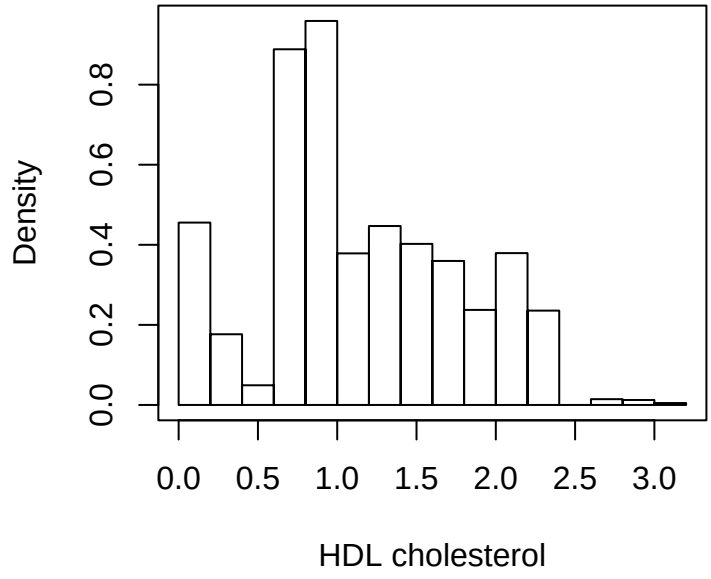
11. Constrained model for non−HDL and HDL: One−dimensional projections

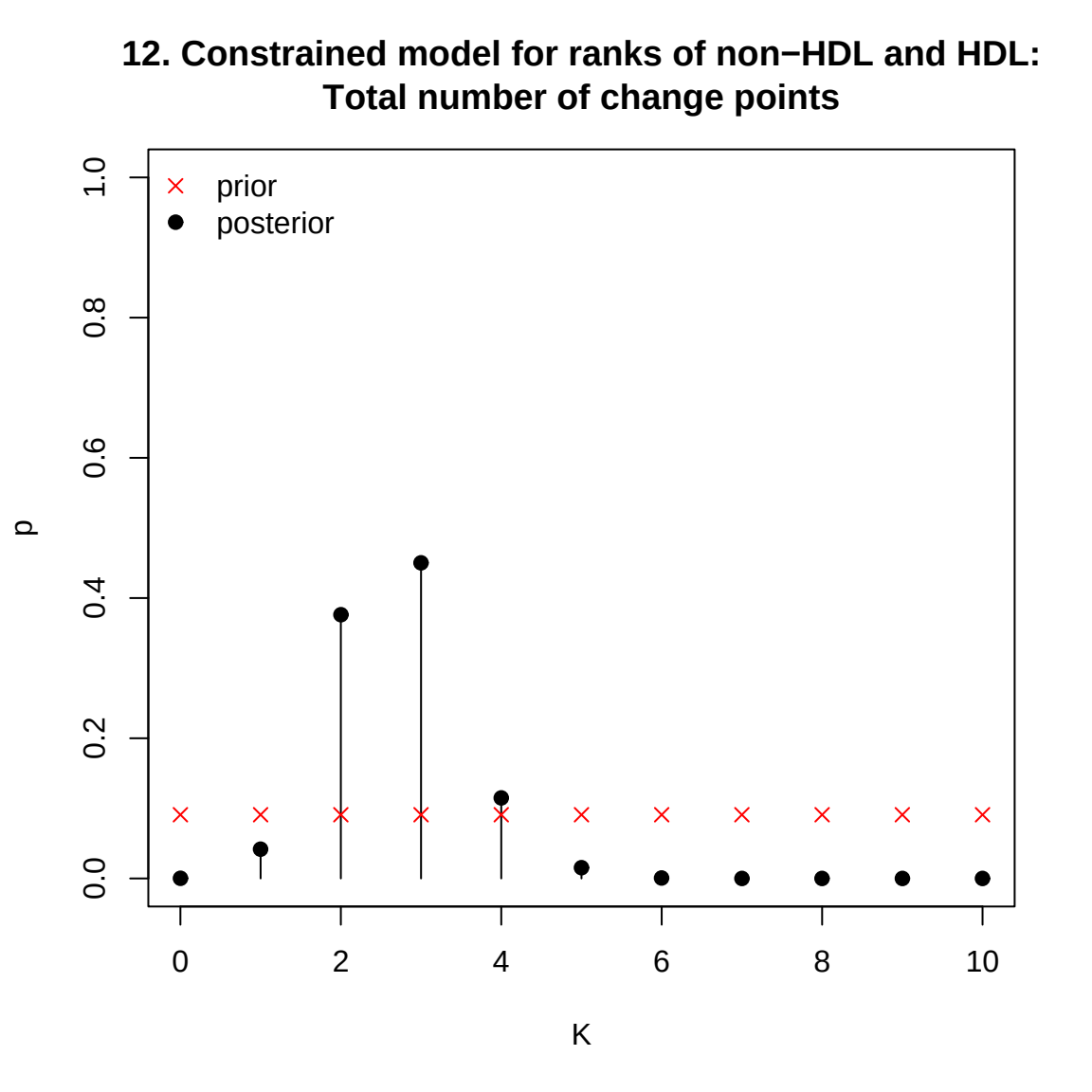




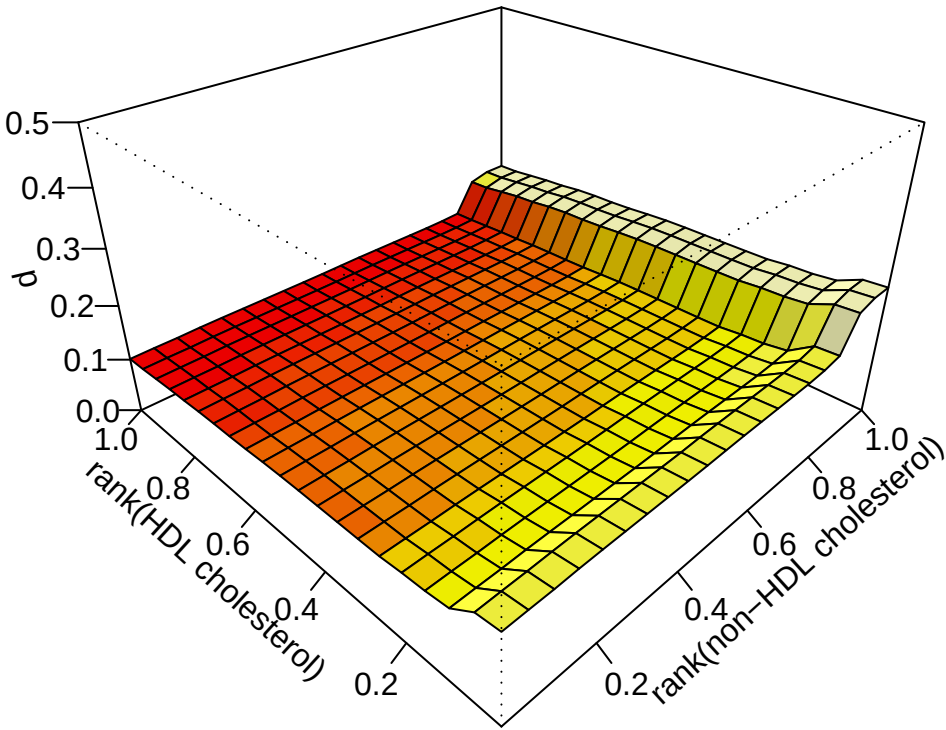
11. Constrained model for non−HDL and HDL: Distributions for change point positions



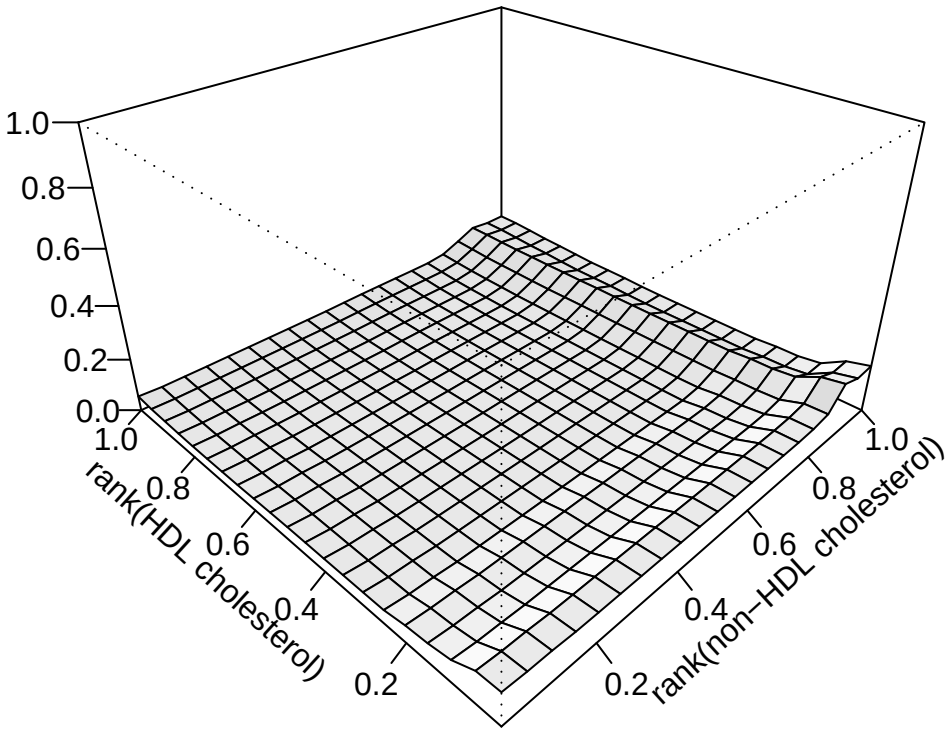



12. Constrained model for ranks of non−HDL and HDL:
Total number of change points
prior
posterior
p
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
K

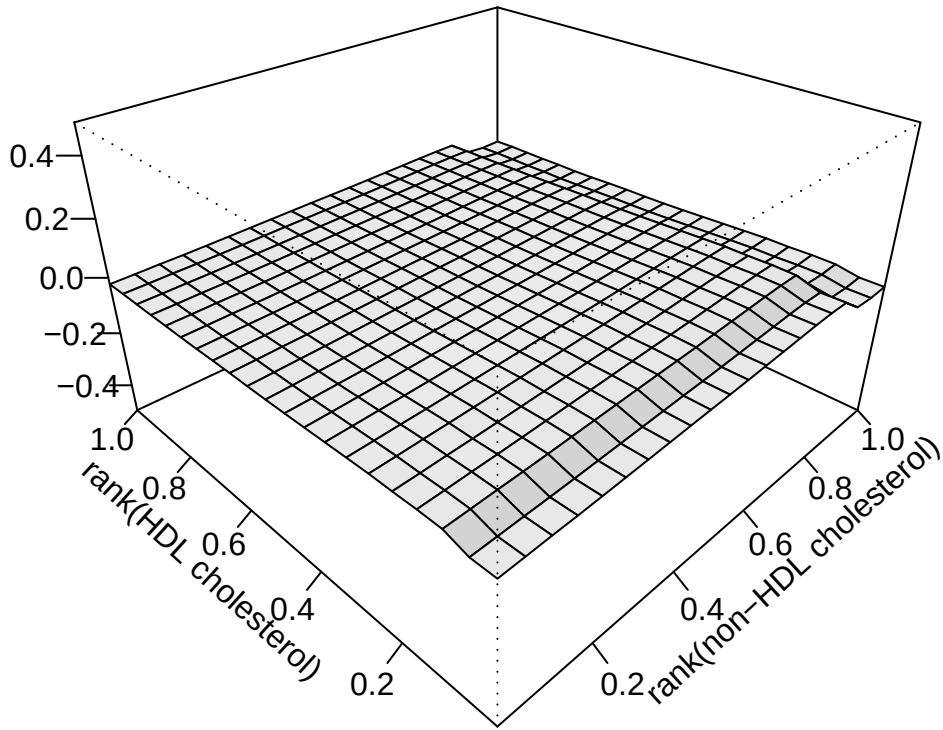
12. Constrained model for ranks of non−HDL and HDL: Posterior median 7 year risk



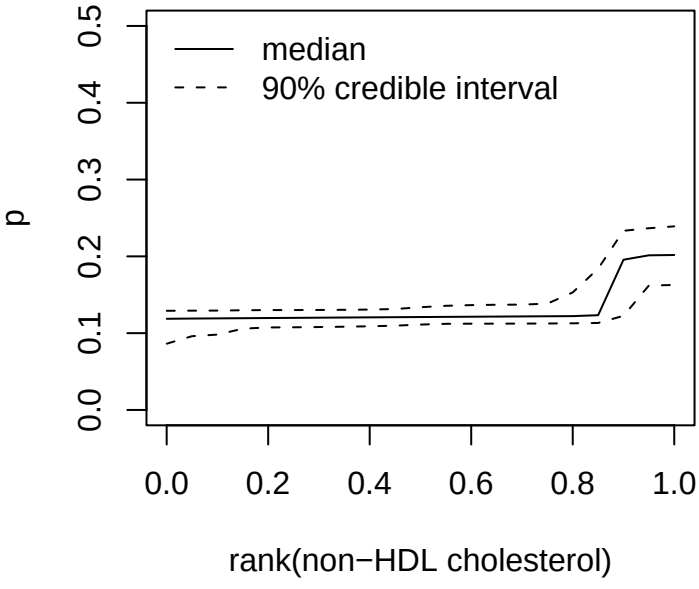
12. Constrained model for ranks of non−HDL and HDL: Length of 90% credible interval

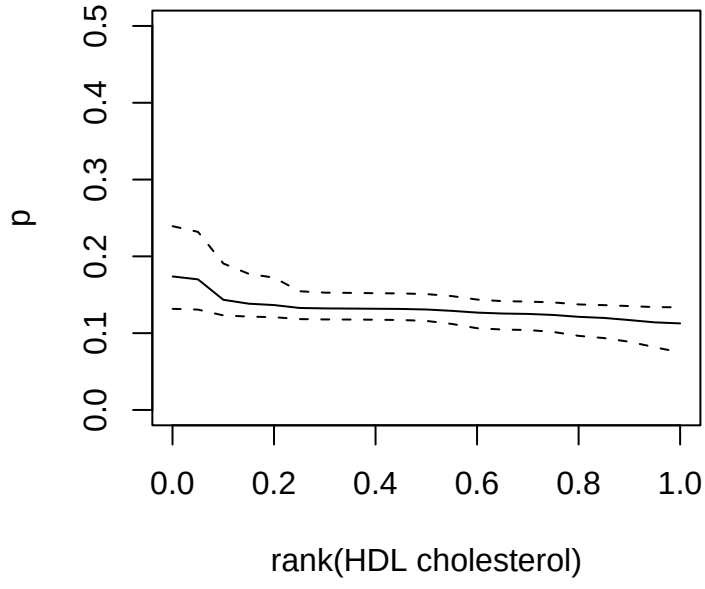


12. Constrained model for ranks of non−HDL and HDL: Difference to proportional hazards model

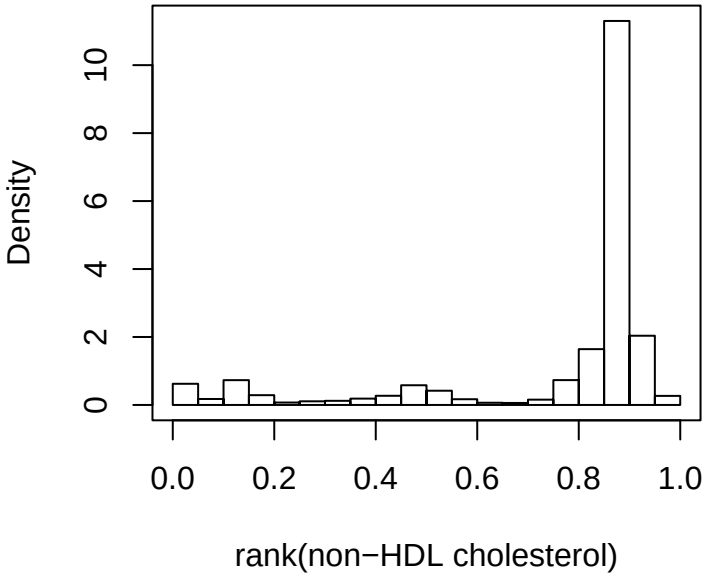


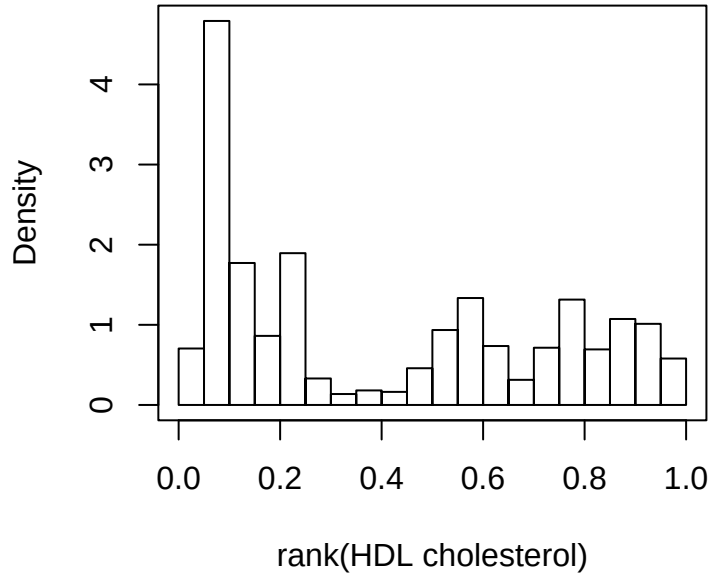
12. Constrained model for ranks of non−HDL and HDL: One−dimensional projections





12. Constrained model for ranks of non−HDL and HDL: Distributions for change point positions





Summary, conclusions

- In the study of biostatistical/epidemiological problems there is a lot of uncritical use of parametric models.
- The resulting parameter estimates may then not ”mean anything real”.
- Structural model assumptions are likely to influence the results of empirical data analysis much more strongly than, e.g., specifying a particular type of prior in a Bayesian model.

Summary, conclusions (2)

- The Bayesian approach to statistical inference adds a great deal of freedom in specifying statistical models for data analysis, by allowing one to think more freely about the structural properties of the models that seem plausible, such as
  - hierarchical or modular structures,
  - natural restrictions in the parameter space,
  - dependencies/conditional independencies,
  - making a distinction between the model describing the underlying state variables and the measurement model,
  - etc.

Summary, conclusions (3)

- The results from data analysis can often be given in terms of probabilities, then involving an integration with respect to the ”unobservables” in the model.
  - This leads to relative robustness of such results with respect to the smoothness properties of the functions being used in the models (”Even very simple models may be O.K.!”).
  - If needed, averaging over non-smooth functions will generally produce smooth reconstructions of the functional relationships that are considered.

Summary, conclusions (4)

- In situations where it can be applied, an assumed monotonicity property of the regression function with respect to covariates has a strong stabilizing effect on the estimates.
- Sometimes mathematically / computationally attractive properties of models are introduced at the cost of unrealistic description of the data. Relevance vs. elegance!
- Careful model diagnostics should be carried out!

References

- Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. *Statistica Sinica*, 4:505-524.
- ——— (1996). Bayesian inference of survival probabilities, under stochastic ordering constraints. *Journal of the American Statistical Association*, 91:1101-1109.
- Dellaportas, P., Forster, J. J. and Ntzoufras, I. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 12:27-36.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711-732.
- Holmes, C. C. and Heard, N. A. (2003). Generalised monotonic regression using random change points. *Statistics in Medicine*, 22:623-638.
- Neelon, B. and Dunson, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics* 60:398-406.
- Schell, M. J. and Singh, B. (1997). The reduced monotonic regression method. *Journal of the American Statistical Association*, 92:128-135.